173500

Requester:
    Jeffrey Russel (TC1600)
Art Unit:
    1654
Employee Number:
    62785
Office Location:
    REM 3D19
Phone_Number:
    571-272-0969
Mailbox Number:
    REM 3C18

Case serial number:
    10/789,494
Class / Subclass(es):
    NA
Earliest Priority Filing Date:
    NA
Format preferred for results:
    Diskette
Search Topic Information:
    Please Search SEQ ID NO:1 (VITTDSDGNE) and SEQ ID NO:5 (YGWGDGGYGSDS)
in STN, in the U.S. patent application sequence databases (pending,
published, and issued), and in Geneseq/Uniprot/Pir. Thank you.
Special Instructions and Other Comments:

RECEIVED
DEC -5 2005
(STIC)

Checked
JRL
12.14.2005

Point of Contact:
Alexandra Waclawiw
Technical Info. Specialist

********************	*********************	********************
Searcher:_____	Type of Search	Vendors and cost where applicable
Searcher Phone:_____ CM1 6A02 Tel 308-449	NA#_____ AA#:___	STN:_____
Date Searcher Picked up:_____ 12-8	S/L:_____ Oligomer:_____	DIALOG:_____
Date completed:_____ 12-12	Encode/Transl:_____	QUESTEL/ORBIT:_____
Searcher Prep Time:_____ 15	Structure #:_____Text:____	LEXIS/NEXIS:_____
Online Time:_____ 15	Inventor:_____ Litigation:____	SEQUENCE SYSTEM:___
		WWW/Internet:_____
		Other (Specify):_____

```
=> d his ful

        (FILE 'HOME' ENTERED AT 08:46:50 ON 12 DEC 2005)

        FILE 'REGISTRY' ENTERED AT 08:46:58 ON 12 DEC 2005
L1              6 SEA ABB=ON   PLU=ON   VITTDSDGNE/SQSP
L2             17 SEA ABB=ON   PLU=ON   YGWGDGGYGSDS/SQSP


        FIlE 'CAPLUS' ENTERED AT 08:47:44 ON 12 DEC 2005
L3              4 SEA ABB=ON   PLU=ON   L1
L4              5 SEA ABB=ON   PLU=ON   L2
```

```
=> fil reg
FILE 'REGISTRY' ENTERED AT 08:48:52 ON 12 DEC 2005
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2005 American Chemical Society (ACS)

Property values tagged with IC are from the ZIC/VINITI data file
provided by InfoChem.

STRUCTURE FILE UPDATES:    11 DEC 2005   HIGHEST RN 869700-38-9
DICTIONARY FILE UPDATES:   11 DEC 2005   HIGHEST RN 869700-38-9

New CAS Information Use Policies, enter HELP USAGETERMS for details.

TSCA INFORMATION NOW CURRENT THROUGH JULY 14, 2005

   Please note that search-term pricing does apply when
   conducting SmartSELECT searches.


*********************************************************************
*                                                                 *
* The CA roles and document type information have been removed from *
* the IDE default display format and the ED field has been added,   *
* effective March 20, 2005.  A new display format, IDERL, is now    *
* available and contains the CA role and document type information. *
*                                                                 , *
*********************************************************************


Structure search iteration limits have been increased.  See HELP SLIMITS
for details.

REGISTRY includes numerically searchable data for experimental and
predicted properties as well as tags indicating availability of
experimental property data in the original document.  For information
on property searching in REGISTRY, refer to:

http://www.cas.org/ONLINE/UG/regprops.html

=> d que l1
L1              6 SEA FILE=REGISTRY ABB=ON   PLU=ON   VITTDSDGNE/SQSP


=> d sqide l1 1-6

L1    ANSWER 1 OF 6  REGISTRY  COPYRIGHT 2005 ACS on STN
RN    803823-75-8  REGISTRY
CN    1: PN: JP2004339189 PAGE: 8 unclaimed sequence (9CI)   (CA INDEX NAME)
FS    PROTEIN SEQUENCE
SQL   151

PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source   |Reference
=========+=============
Not Given|JP2004339189
         |unclaimed
         |PAGE 8
```

```
SEQ       1 MRVKTFVILC CALQYVAYTN ANINDFDEDY FGSDVTVQSS NTTDEIIRDA
         51 SGAVIEEQIT TKKMQRKNKN HGILGKNEKM IKTFVITTDS DGNESIVEED
                                                        ====== ====
        101 VLMKTLSDGT VAQSYVAADA GAYSQSGPYV SNSGYSTHQG YTSDFSTSAA
        151 V
HITS AT:   85-94
MF   Unspecified
CI   MAN
SR   CA
LC   STN Files:   CA, CAPLUS, USPATFULL
DT.CA  CAplus document type:  Patent
RL.P   Roles from patents:  PRP (Properties)
                1 REFERENCES IN FILE CA (1907 TO DATE)
                1 REFERENCES IN FILE CAPLUS (1907 TO DATE)


L1   ANSWER 2 OF 6   REGISTRY   COPYRIGHT 2005 ACS on STN
RN   714954-20-8   REGISTRY
CN   L-Glutamic acid, L-valyl-L-isoleucyl-L-threonyl-L-threonyl-L-α-
     aspartyl-L-seryl-L-α-aspartylglycyl-L-asparaginyl- (9CI)   (CA INDEX
     NAME)
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL  10

SEQ       1 VITTDSDGNE
            ==========
HITS AT:   1-10
MF   C41 H67 N11 O21
SR   CA
LC   STN Files:   CA, CAPLUS, USPATFULL
DT.CA  CAplus document type:  Journal; Patent
RL.P   Roles from patents:  BIOL (Biological study); OCCU (Occurrence); PREP
       (Preparation); USES (Uses)
RL.NP  Roles from non-patents:  BIOL (Biological study); USES (Uses)

Absolute stereochemistry.
```
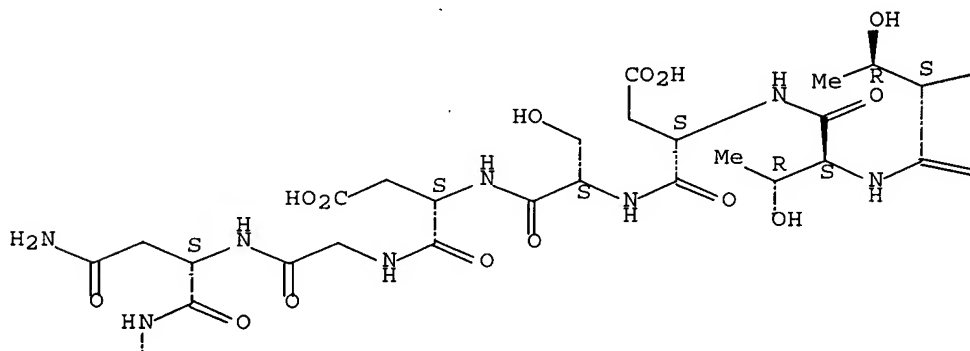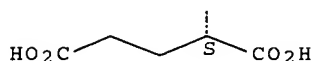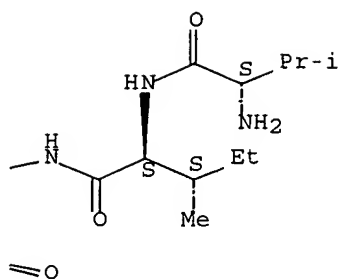
$$\text{O}$$

HN — S Pr-i

NH₂

H N — S Et

O Me

= O

HO₂C — S — CO₂H

PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

2 REFERENCES IN FILE CA (1907 TO DATE)
2 REFERENCES IN FILE CAPLUS (1907 TO DATE)

```
L1    ANSWER 3 OF 6   REGISTRY   COPYRIGHT 2005 ACS on STN
RN    483169-91-1  REGISTRY
CN    GenBank CAA23432 (9CI)    (CA INDEX NAME)
OTHER NAMES:
CN    GenBank CAA23432 (Translated from: GenBank V00094)
FS    PROTEIN SEQUENCE
SQL   168


SEQ      1 MRVKTFVILV CALQYVAYTN ANINDFDEDY FGSDVTVQSS NTTDEIIRDA
        51 SGAVIEEQIT TKKMQRKNKN HGILGKNEKM IKTFVITTDS DGNESIVEED
                                                     ====== ====
       101 VLMKTLSDGT VAQSYVAADA GAYSQSGPYV SNSGYSTHQG YTSDFSTSAA
       151 VGAGAGAGAA AGSGAGAG
HITS AT:    85-94
MF    Unspecified
CI    MAN
SR    GenBank


L1    ANSWER 4 OF 6   REGISTRY   COPYRIGHT 2005 ACS on STN
RN    482255-67-4  REGISTRY
CN    GenBank AAA27838 (9CI)    (CA INDEX NAME)
OTHER NAMES:
CN    GenBank AAA27838 (Translated from: GenBank M24222)
FS    PROTEIN SEQUENCE
SQL   178


SEQ      1 MRVKTFVILC CALQYVAYTN ANINDFDEDY FGSDVTVQSS NTTDEIIRDA
        51 SGAVIEEEIT TKKMQRKNKN HGILGKNEKM IKTFVITTDS DGNESIVEED
                                                     ====== ====
       101 VLMKTLSDGT VAQSYVAADA GAYSQSGPYV SNSGYSTHQG YRSDFASAAV
       151 GAGAGAGAAA GSGAGAGAGY GAASGAGA
```

```
HITS AT:    85-94
MF    Unspecified
CI    MAN
SR    GenBank


L1    ANSWER 5 OF 6   REGISTRY   COPYRIGHT 2005 ACS on STN
RN    482246-94-6  REGISTRY
CN    GenBank CAA27612 (9CI)   (CA INDEX NAME)
OTHER NAMES:
CN    GenBank CAA27612 (Translated from: GenBank X03973)
FS    PROTEIN SEQUENCE
SQL   178


SEQ       1 MRVKTFVILV CALQYVAYTN ANINDFDEDY FGSDVTVQSS NTTDEIIRDA
         51 SGAVIEEEIT TKKMQRKNKN HGILGKNEKM IKTFVITTDS DGNESIVEED
                                                      ====== ====
        101 VLMKTLSDGT VAQSYVAADA GAYSQSGPYV SNSGYSTHQG YRSDFASAAV
        151 GAGAGAGAAA GSGAGAGAGY GAASGAGA
HITS AT:    85-94
MF    Unspecified
CI    MAN
SR    GenBank


L1    ANSWER 6 OF 6   REGISTRY   COPYRIGHT 2005 ACS on STN
RN    303229-60-9  REGISTRY
CN    Fibroin (silkworm strain p50 heavy chain) (9CI)   (CA INDEX NAME)
OTHER NAMES:
CN    Fibroin (Bombyx mori strain p50 gene fib-H heavy chain)
CN    GenBank AAF76983
CN    GenBank AAF76983 (Translated from: GenBank AF226688)
FS    PROTEIN SEQUENCE
SQL   5263


SEQ       1 MRVKTFVILC CALQYVAYTN ANINDFDEDY FGSDVTVQSS NTTDEIIRDA
         51 SGAVIEEQIT TKKMQRKNKN HGILGKNEKM IKTFVITTDS DGNESIVEED
                                                      ====== ====
        101 VLMKTLSDGT VAQSYVAADA GAYSQSGPYV SNSGYSTHQG YTSDFSTSAA
        151 VGAGAGAGAA AGSGAGAGAG YGAASGAGAG AGAGAGAGYG TGAGAGAGAG
        201 YGAGAGAGAG AGYGAGAGAG AGAGYGAGAG AGAGAGYGAG AGAGAGAGYG
        251 AGAGAGAGAG YGAASGAGAG AGYGQGVGSG AASGAGAGAG AGSAAGSGAG
        301 AGAGTGAGAG YGAGAGAGAG AGYGAASGTG AGYGAGAGAG YGGASGAGAG
        351 AGAGAGAGAG AGYGTGAGYG AGAGAGAGAG AGAGYGAGAG AGYGAGYGVG
        401 AGAGYGAGYG AGAGSGAASG AGSGAGAGSG AGAGSGAGAG SGAGAGSGAG
        451 AGSGAGAGSG AGAGSGAGAG SGTGAGSGAG AGYGAGAGAG YGAGAGSGAA
        501 SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG SGAGAGYGAG AGAGYGAGAG
        551 AGYGAGAGVG YGAGAGSGAA SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG
        601 SGAGAGSGAG AGSGAGAGSG AGAGSGAGVG YGAGVGAGYG AGYGAGAGAG
        651 YGAGAGSGAA SGAGAGAGAG AGTGSSGFGP YVANGGYSRS DGYEYAWSSD
        701 FGTGSGAGAG SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG YGAGVGVGYG
        751 AGYGAGAGAG YGAGAGSGAA SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG
        801 SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG SGAGVGSGAG AGSGAGAGVG
        851 YGAGAGVGYG AGAGSGAASG AGAGSGAGAG SGAGAGSGAG AGSGAGAGSG
        901 AGAGSGAGAG SGAGAGSGAG AGSGAGVGYG AGVGAGYGAG YGAGAGAGYG
        951 AGAGSGAASG AGAGSGAGAG SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG
       1001 SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG YGAGAGAGYG AGYGAGAGAG
       1051 YGAGAGSGAA SGAGSGAGAG SGAGAGAGSG AGAGSGAGAG SGAGAGSGAG
       1101 AGSGAGAGSG AGAGYGAGVG AGYGAGYGAG AGAGYGAGAG SGAASGAGAG
       1151 SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG SGAGAGSGAG VGYGAGYGAG
       1201 AGAGYGAGAG SGAASGAGAG AGAGAGTGSS GFGPYVAHGG YSGYEYAWSS
       1251 ESDFGTGSGA GAGSGAGAGS GAGAGSGAGA GSGAGYGAGV GAGYGAGYGA
```

```
1301 GAGAGYGAGA GSGAGSGAGA GSGAGAGSGA GAGSGAGAGS GAGAGSGAGA
1351 GSGAGAGSGA GAGSGAGAGY GAGYGAGAGA GYGAGAGSGA GSGAGAGSGA
1401 GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA GAGSGAGAGY GAGVGAGYGA
1451 GYGAGAGAGY GAGAGSGAGS GAGAGSGAGA GSGAGAGSGA GVGSGAGAGS
1501 GAGAGSGAGA GSGAGAGYGA GYGAGAGAGY GAGAGSGAGS GAGAGSGAGA
1551 GSGAGAGSGA GAGSGAGAGS GAGAGSGAGA GSGAGVGYGA GVGAGYGAGY
1601 GAGAGAGYGA GAGSGAASGA GAGAGAGAGT GSSGFGPYVA NGGYSGYEYA
1651 WSSESDFGTG SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG YGAGYGAGAG
1701 AGYGAGAGSG AGSGAGAGSG AGAGSGAGAG SGAGAGSGAG AGSGAGAGSG
1751 AGAGSGAGSG SGAGAGSGAG AGSGAGAGYG AGVGAGYGVG YGAGAGAGYG
1801 AGAGSGAASG AGAGAGAGAG TGSSGFGPYV AHGGYSGYEY AWSSESDFGT
1851 GSGAGAGSGA GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA GAGYGAGVGA
1901 GYGAAYGAGA GAGYGAGAGS GAASGAGAGS GAGAGSGAGA GSGAGAGSGA
1951 GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA GAGSGAGAGY GAGAGAGYGA
2001 GAGSGAGSGA GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA GSGSGAGAGS
2051 GAGAGSGAGA GYGAGVGAGY GAGYGAGAGA GYGAGAGSGA GSGAGAGSGA
2101 GAGYGAGAGA GYGAGYGAGA GAGYGAGAGT GAGSGAGAGS GAGAGSGAGA
2151 GSGAGAGSGA GAGSGAGAGS GAGSGSGAGA GSGAGAGSGA GAGSGAGAGS
2201 GAGAGSGAGA GYGAGAGAGY GAGYGAGAGA GYGAGAGSGA GSGAGAGSGA
2251 GAGSGAGAGS GAGAGYGAGY GAGAGSGAAS GAGAGAGAGA GTGSSGFGPY
2301 VAHGGYSGYE YAWSSESDFG TGSGAGAGSG AGAGAGAGAG SGAGAGYGAG
2351 VGAGYGAGYG AGAGAGYGAG AGSGTGSGAG AGSGAGAGYG AGVGAGYGAG
2401 AGSGAAFGAG AGAGAGSGAG AGSGAGAGSG AGAGSGAGAG SGAGAGYGAG
2451 YGAGVGAGYG AGAGSGAASG AGAGSGAGAG SGAGAGSGAG AGSGAGAGSG
2501 AGAGYGAGVG AGYGAGYGAG AGAGYGAGAG SGAASGAGAG SGAGAGAGSG
2551 AGAGSGAGAG SGAGAGSGAG SGAGAGSGAG AGSGAGAGYG AGAGSGAASG
2601 AGAGAGAGAG TGSSGFGPYV ANGGYSGYEY AWSSESDFGT GSGAGAGSGA
2651 GAGSGAGAGS GAGAGSGAGA GYGAGVGAGY GAGYGAGAGA GYGAGAGSGA
2701 GSGAGAGSGA GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA GAGYGAGAGS
2751 GAASGAGAGS GAGAGSGAGA GSGAGAGSGA GAGSGAGAGS GAGAGYGAGV
2801 GAGYGVGYGA GAGAGYGAGA GSGAGSGAGA GSGAGAGSGA GAGSGAGAGS
2851 GAGSGAGAGS GAGAGSGAGA GSGAGSGAGA GSGAGAGYGV GYGAGAGAGY
2901 GAGAGSGAGS GAGAGSGAGA GSGAGAGSGA GSGAGAGSGA GAGSGAGAGS
2951 GAGAGYGAGV GAGYGVGYGA GAGAGYGAGA GSGAGSGAGA GSGAGAGSGA
3001 GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA GSGAGAGSGA GAGSGAGAGS
3051 GAGAGSGAGS GAGAGSGAGA GSGAGAGSGA GAGYGAGVGA GYGVGYGAGV
3101 GAGYGAGAGS GAASGAGAGS GAGAGAGSGA GAGSGAGAGS GAGAGSGAGA
3151 GSGAGAGSGA GAGYGAGYGA GVGAGYGAGA GVGYGAGAGA GYGAGAGSGA
3201 ASGAGAGAGS GAGAGTGAGA GSGAGAGYGA GAGSGAASGA GAGAGAGAGT
3251 GSSGFGPYVA NGGYSGYEYA WSSESDFGTG SGAGAGSGAG AGSGAGAGSG
3301 AGAGSGAGAG YGAGVGAGYG AGAGSGAGSG AGAGSGAGAG SGAGAGSGAG
3351 AGSGAGAGYG AGAGSGTGSG AGAGSGAGAG SGAGAGSGAG AGSGAGAGSG
3401 AGAGSGVGAG YGVGYGAGAG AGYGVGYGAG AGAGYGAGAG SGTGSGAGAG
3451 SGAGAGSGAG AGSGAGAGSG AGAGSGAGAG SGAGAGYGAG VGAGYGVGYG
3501 AGAGAGYGAG AGSGAGSGAG AGSGAGAGSG AGAGSGAGAG SGAGSGAGAG
3551 SGAGAGSGAG AGSGAGSGAG AGSGAGAGYG VGYGAGAGAG YGAGAGSGAG
3601 SGAGAGSGAG AGSGAGAGSG AGSGAGAGSG AGAGSGAGAG SGAGAGYGAG
3651 VGAGYGVGYG AGAGAGYGAG AGSGAGSGAG AGSGAGAGSG AGAGSGAGAG
3701 SGAGAGSGAG AGSGAGAGSG AGSGAGAGSG AGAGSGAGAG SGAGAGYGAG
3751 VGAGYGVGYG AGAGAGYGAG AGSGAASGAG AGAGAGAGTG SSGFGPYVAN
3801 GGYSGYEYAW SSESDFGTGS GAGAGSGAGA GSGAGAGYGA GYGAGVGAGY
3851 GAGAGVGYGA GAGAGYGAGA GSGAASGAGA GAGAGAGSGA GAGSGAGAGA
3901 GSGAGAGYGA GYGIGVGAGY GAGAGVGYGA GAGAGYGAGA GSGAASGAGA
3951 GSGAGAGSGA GAGSGAGAGS GAGAGSGAGA GSGAGAGYGA GYGAGVGAGY
4001 GAGAGVGYGA GAGAGYGAGA GSGAASGAGA GAGAGAGAGS GAGAGSGAGA
4051 GSGAGAGSGA GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA GAGYGAGVGA
4101 GYGAGYGGAG AGYGAGAGSG AASGAGAGSG AGAGSGAGAG SGAGAGSGAG
4151 AGSGAGAGYG AGAGSGAASG AGAGAGAGAG TGSSGFGPYV NGGYSGYEYA
4201 WSSESDFGTG SGAGAGSGAG AGSGAGAGYG AGVGAGYGAG YGAGAGAGYG
4251 AGAGSGAASG AGAGSGAGAG SGAGAGSGAG AGSGAGSGAG AGSGAGAGSG
```

```
4301 AGAGSGAGAG SGAGAGSGAG AGYGAGVGAG YGAGYGAGAG AGYGAGAGSG
4351 AASGAGAGSG AGAGAGSGAG AGSGAGAGSG AGAGSGAGAG SGAGAGSGAG
4401 SGAGAGSGAG AGYGAGYGAG VGAGYGAGAG VGYGAGAGAG YGAGAGSGAA
4451 SGAGAGSGSG AGSGAGAGSG AGAGSGAGAG AGSGAGAGSG AGAGSGAGAG
4501 YGAGYGAGAG SGAASGAGAG AGAGAGTGSS GFGPYVANGG YSGYEYAWSS
4551 ESDFGTGSGA GAGSGAGAGS GAGAGYGAGV GAGYGAGYGA GAGAGYGAGA
4601 GSGAGSGAGA GSGAGAGSGA GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA
4651 GAGYGAGYGA GAGAGYGAGA GVGYGAGAGA GYGAGAGSGA GSGAGAGSGS
4701 GAGAGSGSGA GSGAGAGSGA GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA
4751 GAGYGAGYGI GVGAGYGAGA GVGYGAGAGA GYGAGAGSGA ASGAGAGSGA
4801 GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA GAGSGAGAGS GAGAGSGAGA
4851 GYGAGAGVGY GAGAGSGAAS GAGAGSGAGA GSGAGAGSGA GAGSGAGAGS
4901 GAGAGSGAGA GSGAGSGAGA GSGAGAGYGA GYGAGVGAGY GAGAGYGAGY
4951 GVGAGAGYGA GAGSGAGSGA GAGSGAGAGS GAGAGSGAGA GSGAGAGSGA
5001 GSGAGAGYGA GAGAGYGAGA GAGYGAGAGS GAASGAGAGA GAGSGAGAGS
5051 GAGAGSGAGS GAGAGSGAGA GYGAGAGSGA ASGAGAGSGA GAGAGAGAGA
5101 GSGAGAGSGA GAGYGAGAGS GAASGAGAGA GAGTGSSGFG PYVANGGYSR
5151 REGYEYAWSS KSDFETGSGA ASGAGAGAGS GAGAGSGAGA GSGAGAGSGA
5201 GAGGSVSYGA GRGYGQGAGS AASSVSSASS RSYDYSRRNV RKNCGIPRRQ
5251 LVVKFRALPC VNC
HITS AT:    85-94
MF    Unspecified
CI    MAN
SR    CA
LC    STN Files:    CA, CAPLUS
DT.CA CAplus document type:   Journal
RL.NP  Roles from non-patents:   BIOL (Biological study);  PRP (Properties)
              2 REFERENCES IN FILE CA (1907 TO DATE)
              2 REFERENCES IN FILE CAPLUS (1907 TO DATE)

                                          .


=> d sqide l2 1-17

L2    ANSWER 1 OF 17   REGISTRY   COPYRIGHT 2005 ACS on STN
RN    803823-78-1  REGISTRY
CN    L-Serine, L-serylglycyl-L-phenylalanyl-L-tyrosyl-L-α-glutamyl-L-
      threonyl-L-histidyl-L-α-aspartyl-L-seryl-L-tyrosyl-L-seryl-L-seryl-L-
      tyrosylglycyl-L-serylglycyl-L-seryl-L-seryl-L-seryl-L-alanyl-L-alanyl-L-
      alanyl-L-alanyl-L-seryl-L-serylglycyl-L-alanylglycylglycyl-L-
      alanylglycylglycylglycyl-L-tyrosylglycyl-L-tryptophylglycyl-L-α-
      aspartylglycylglycyl-L-tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI)
      (CA INDEX NAME)
OTHER NAMES:
CN    13: PN: JP2004339189 PAGE: 9 unclaimed sequence
FS    PROTEIN SEQUENCE
SQL   45

PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source   |Reference
=========+=============
Not Given|JP2004339189
         |unclaimed
         |PAGE 9


    .

SEQ     1 SGFYETHDSY SSYGSGSSSA AAASSGAGGA GGGYGWGDGG YGSDS

                                   ======= =====
HITS AT:    34-45
```

```
MF    Unspecified
CI    MAN
SR    CA
LC    STN Files:   CA, CAPLUS, USPATFULL
DT.CA CAplus document type:  Patent
RL.P  Roles from patents:  PRP (Properties)
               1 REFERENCES IN FILE CA (1907 TO DATE)
               1 REFERENCES IN FILE CAPLUS (1907 TO DATE)


L2    ANSWER 2 OF 17  REGISTRY  COPYRIGHT 2005 ACS on STN
RN    803731-54-6  REGISTRY
CN    L-Serine, glycyl-L-serylglycyl-L-alanylglycylglycyl-L-
      serylglycylglycylglycyl-L-tyrosylglycyl-L-tryptophylglycyl-L-α-
      aspartylglycylglycyl-L-tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI)
      (CA INDEX NAME)
OTHER NAMES:
CN    44: PN: JP2004339189 PAGE: 10 unclaimed sequence
FS    PROTEIN SEQUENCE; STEREOSEARCH
SQL   22


PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source   |Reference
=========+=============
Not Given|JP2004339189
         |unclaimed
         |PAGE 10



SEQ      1 GSGAGGSGGG YGWGDGGYGS DS
           ========== ==
HITS AT:    11-22
MF    C76 H101 N23 O33
SR    CA
LC    STN Files:   CA, CAPLUS, USPATFULL
DT.CA CAplus document type:  Patent
RL.P  Roles from patents:  PRP (Properties)

Absolute stereochemistry.
```

PAGE 1-A

PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

        1 REFERENCES IN FILE CA (1907 TO DATE)
        1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

L2   ANSWER 3 OF 17   REGISTRY   COPYRIGHT 2005 ACS on STN
RN   803731-44-4   REGISTRY
CN   L-Serine, L-serylglycyl-L-alanylglycylglycyl-L-threonylglycylglycylglycyl-
     L-tyrosylglycyl-L-tryptophylglycyl-L-α-aspartylglycylglycyl-L-

tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI)   (CA INDEX NAME)
OTHER NAMES:
CN   35: PN: JP2004339189 PAGE: 10 unclaimed sequence
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL 21

PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source |Reference
=========+============
Not Given|JP2004339189
         |unclaimed
         |PAGE 10

SEQ   1 SGAGGTGGGY GWGDGGYGSD S
           = ========== =
HITS AT: 10-21
MF   C75 H100 N22 O32
SR   CA
LC   STN Files: CA, CAPLUS, USPATFULL
DT.CA CAplus document type: Patent
RL.P   Roles from patents: PRP (Properties)

Absolute stereochemistry.

                                                        PAGE 1-A

PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

                    1 REFERENCES IN FILE CA (1907 TO DATE)
                    1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

L2   ANSWER 4 OF 17   REGISTRY   COPYRIGHT 2005 ACS on STN
RN   803731-40-0 REGISTRY
CN   L-Serine, glycyl-L-serylglycyl-L-alanylglycylglycyl-L-
      alanylglycylglycylglycyl-L-tyrosylglycyl-L-tryptophylglycyl-L-α-
      aspartylglycylglycyl-L-tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI)
      (CA INDEX NAME)
OTHER NAMES:
CN   33: PN: JP2004339189 PAGE: 10 unclaimed sequence
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL   22

PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source |Reference
=========+============
Not Given|JP2004339189
         |unclaimed
         |PAGE 10

SEQ   1 GSGAGGAGGG YGWGDGGYGS DS
            ========== ==
HITS AT: 11-22
MF   C76 H101 N23 O32
SR   CA
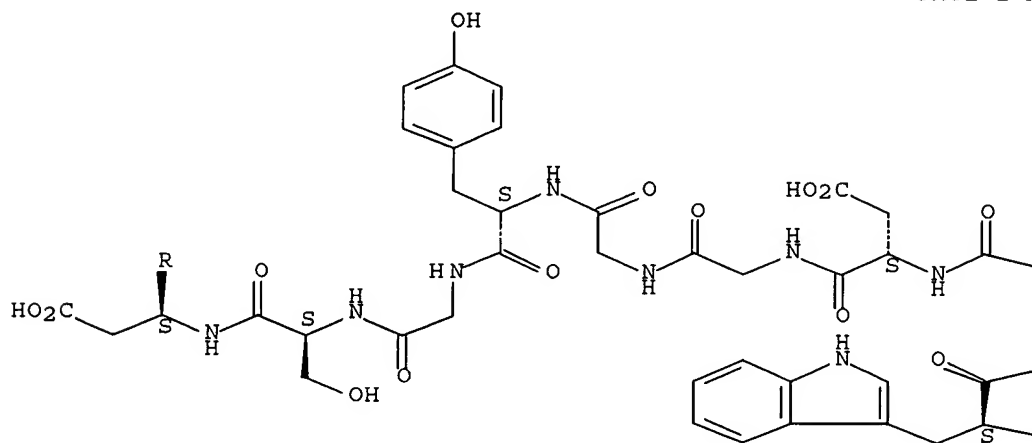LC   STN Files: CA, CAPLUS, USPATFULL
DT.CA CAplus document type: Patent
RL.P Roles from patents: PRP (Properties)

Absolute stereochemistry.
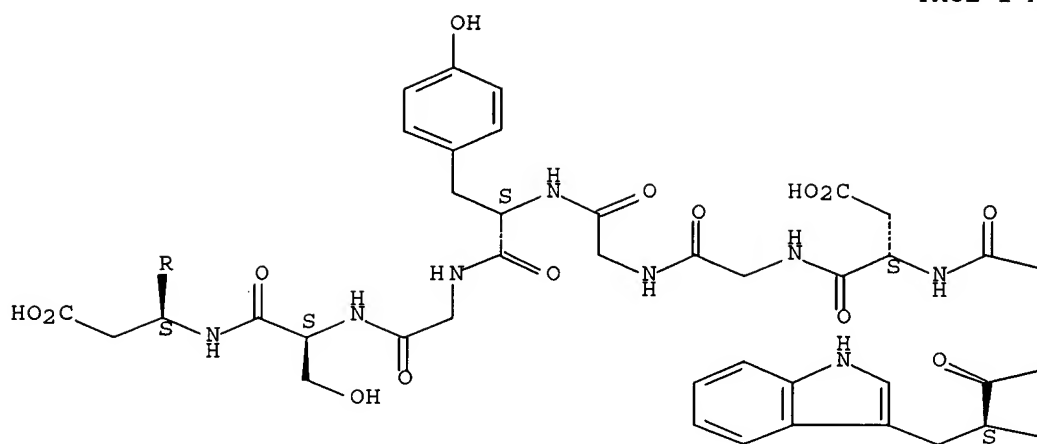
PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

         1 REFERENCES IN FILE CA (1907 TO DATE)
         1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

L2   ANSWER 5 OF 17   REGISTRY   COPYRIGHT 2005 ACS on STN
RN   803731-38-6 REGISTRY
CN   L-Serine, glycyl-L-serylglycyl-L-alanylglycylglycyl-L-
      valylglycylglycylglycyl-L-tyrosylglycyl-L-tryptophylglycyl-L-α-

aspartylglycylglycyl-L-tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI)
(CA INDEX NAME)
OTHER NAMES:
CN   28: PN: JP2004339189 PAGE: 9 unclaimed sequence
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL   22

PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source |Reference
=========+=============
Not Given|JP2004339189
         |unclaimed
         |PAGE 9

SEQ   1 GSGAGGVGGG YGWGDGGYGS DS
          ========== ==
HITS AT: 11-22
MF   C78 H105 N23 O32
SR   CA
LC   STN Files: CA, CAPLUS, USPATFULL
DT.CA CAplus document type: Patent
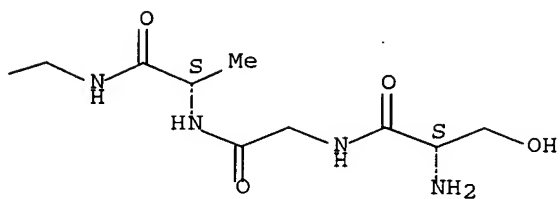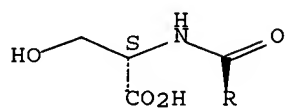RL.P   Roles from patents: PRP (Properties)
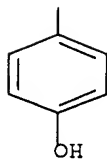
Absolute stereochemistry.
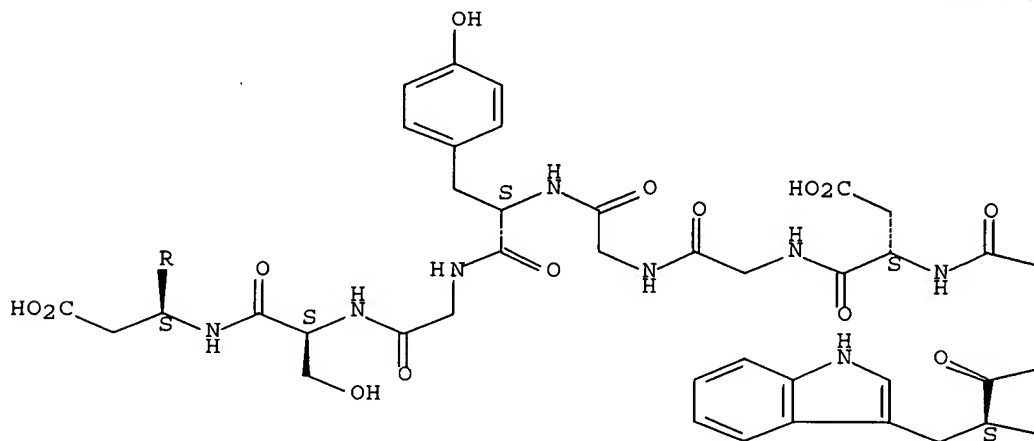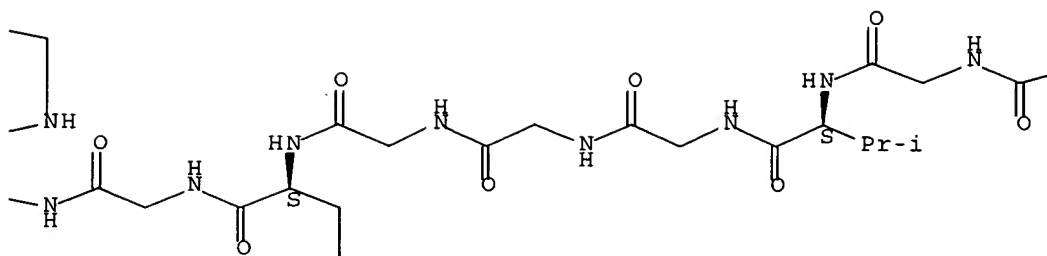
PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

                1 REFERENCES IN FILE CA (1907 TO DATE)
                1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

L2   ANSWER 6 OF 17   REGISTRY   COPYRIGHT 2005 ACS on STN
RN   803731-34-2   REGISTRY
CN   L-Serine, L-serylglycyl-L-alanylglycylglycyl-L-alanylglycylglycylglycyl-L-
      tyrosylglycyl-L-tryptophylglycyl-L-α-aspartylglycylglycyl-L-
      tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI)   (CA INDEX NAME)
OTHER NAMES:
CN   25: PN: JP2004339189 PAGE: 9 unclaimed sequence
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL   21

PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source |Reference
=========+============
Not Given|JP2004339189
         |unclaimed
         |PAGE 9

SEQ   1 SGAGGAGGGY GWGDGGYGSD S
              = ========== =
HITS AT: 10-21
MF   C74 H98 N22 O31
SR   CA
LC   STN Files: CA, CAPLUS, USPATFULL
DT.CA CAplus document type: Patent
RL.P Roles from patents: PRP (Properties)

Absolute stereochemistry.
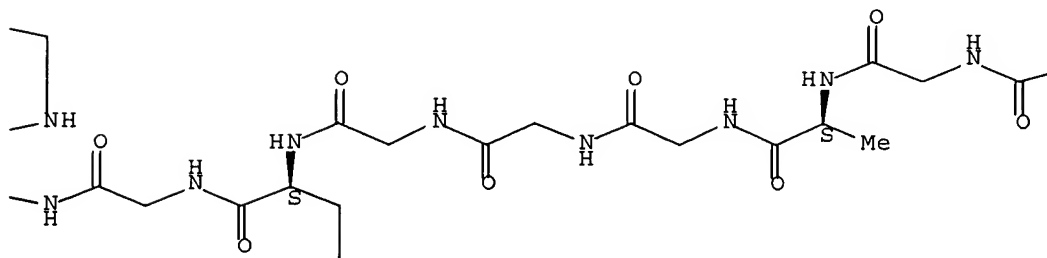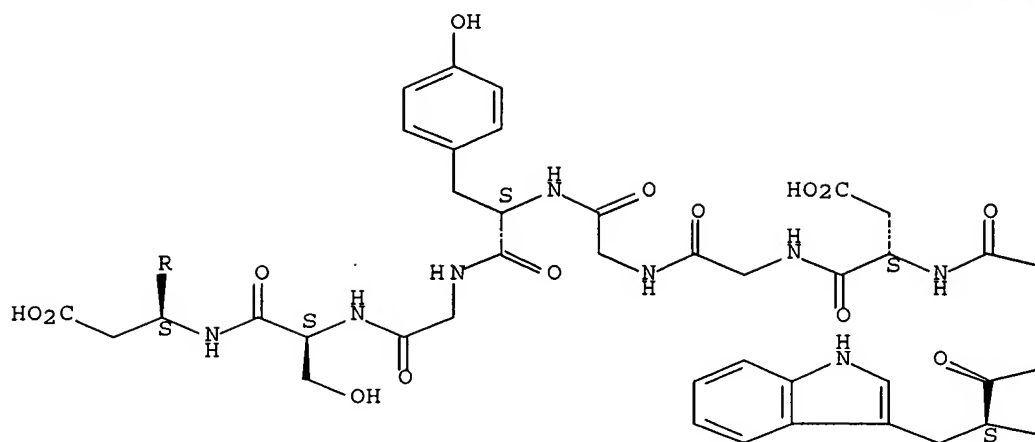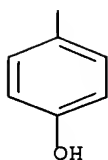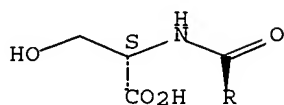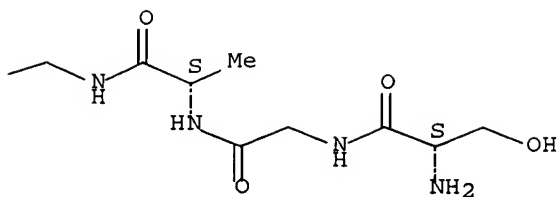
PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

       1 REFERENCES IN FILE CA (1907 TO DATE)
       1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

L2   ANSWER 7 OF 17 REGISTRY COPYRIGHT 2005 ACS on STN
RN   803731-33-1 REGISTRY
CN   L-Serine, glycyl-L-serylglycyl-L-alanylglycylglycyl-L-alanylglycylglycyl-L-
     α-aspartyl-L-tyrosylglycyl-L-tryptophylglycyl-L-α-
     aspartylglycylglycyl-L-tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI)
     (CA INDEX NAME)
OTHER NAMES:

```
CN    24: PN: JP2004339189 PAGE: 9 unclaimed sequence
FS    PROTEIN SEQUENCE; STEREOSEARCH
SQL   22
```

PATENT ANNOTATIONS (PNTE):

```
Sequence  |Patent
Source    |Reference
==========+=============
Not Given |JP2004339189
          |unclaimed
          |PAGE 9
```

```
SEQ      1 GSGAGGAGGD YGWGDGGYGS DS
                    ========== ==
HITS AT:    11-22
MF    C78 H103 N23 O34
SR    CA
LC    STN Files:   CA, CAPLUS, USPATFULL
DT.CA CAplus document type:  Patent
RL.P   Roles from patents:  PRP (Properties)
```
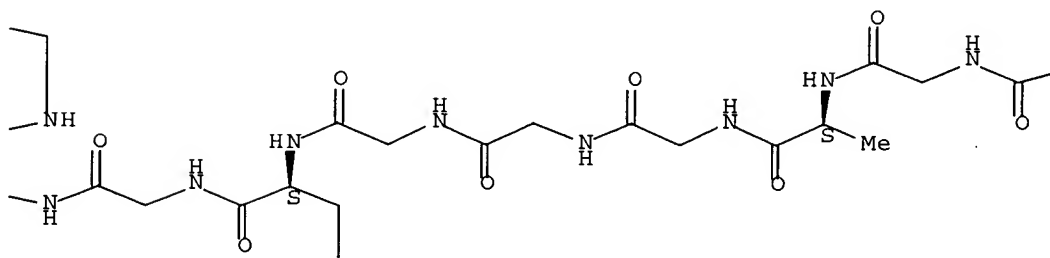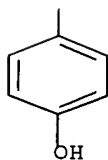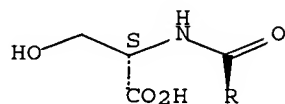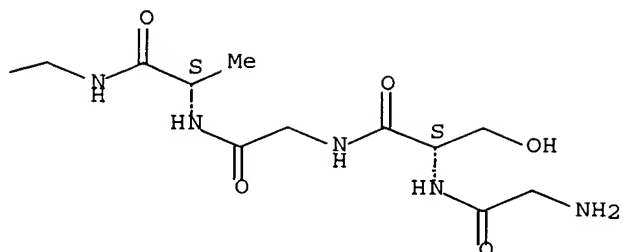
Absolute stereochemistry.
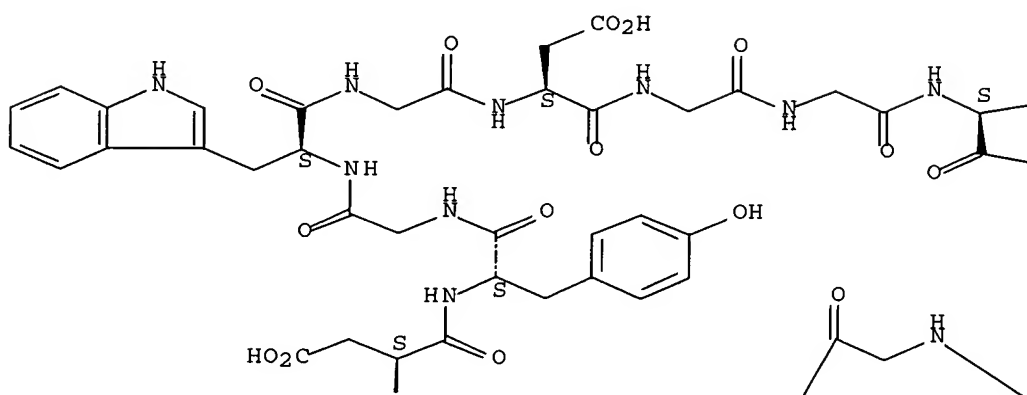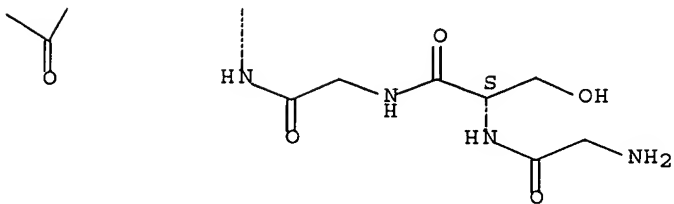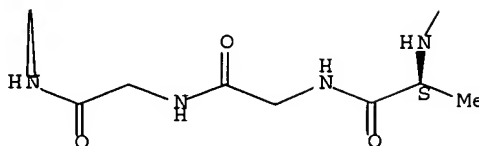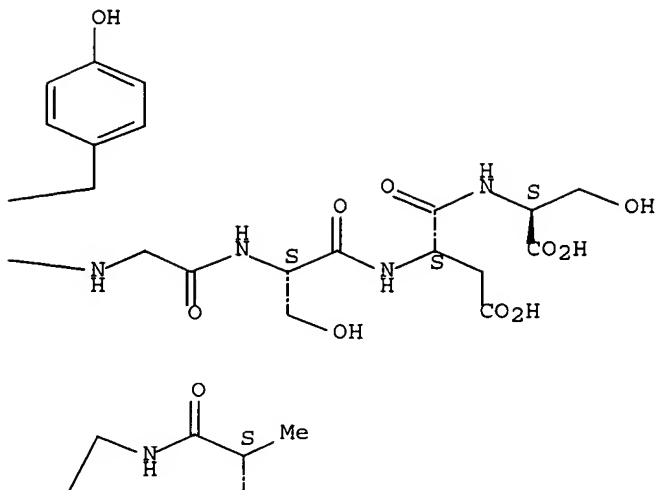
PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

        1 REFERENCES IN FILE CA (1907 TO DATE)
        1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

L2   ANSWER 8 OF 17   REGISTRY   COPYRIGHT 2005 ACS on STN
RN   803731-31-9   REGISTRY
CN   L-Serine, L-serylglycyl-L-alanylglycylglycyl-L-serylglycyl-L-arginylglycyl-
      L-tyrosylglycyl-L-tryptophylglycyl-L-α-aspartylglycylglycyl-L-
      tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI)   (CA INDEX NAME)
OTHER NAMES:
CN   22: PN: JP2004339189 PAGE: 9 unclaimed sequence
FS   PROTEIN SEQUENCE; STEREOSEARCH

SQL   21

SEQ   1 SGAGGSGRGY GWGDGGYGSD S
           = ========== =
HITS AT: 10-21
MF   C78 H107 N25 O32
SR   CA
LC   STN Files: CA, CAPLUS, USPATFULL
DT.CA CAplus document type: Patent
RL.P   Roles from patents: PRP (Properties)

Absolute stereochemistry.
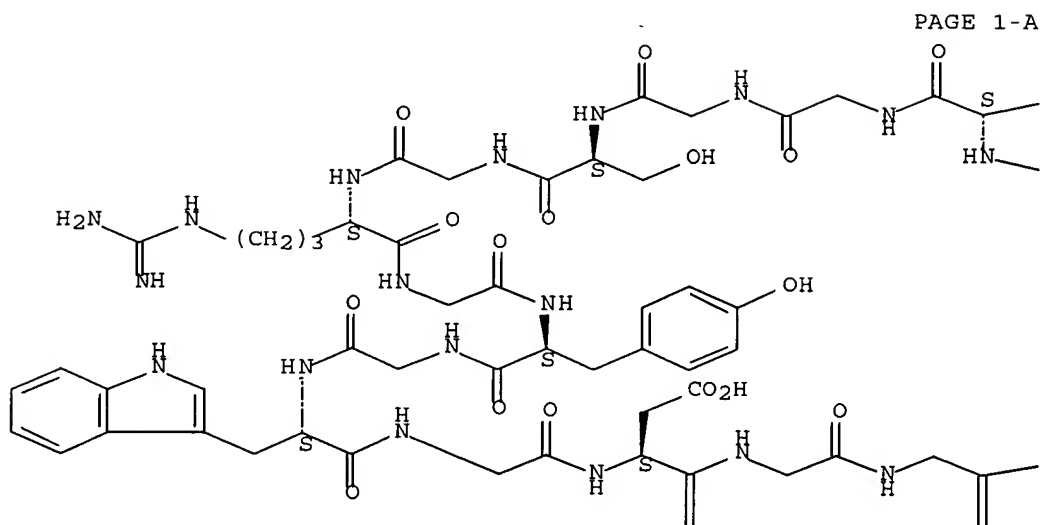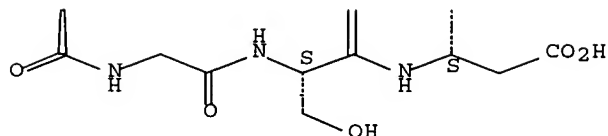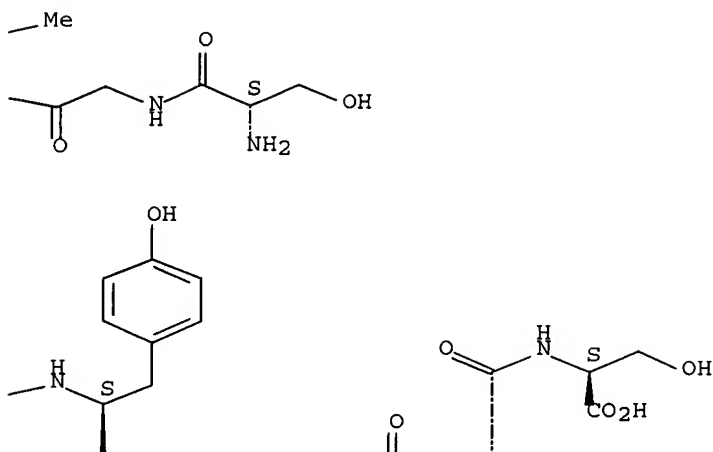
PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

          1 REFERENCES IN FILE CA (1907 TO DATE)
          1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

L2   ANSWER 9 OF 17 REGISTRY   COPYRIGHT 2005 ACS on STN
RN   803731-29-5 REGISTRY
CN   L-Serine, glycyl-L-α-aspartyl-L-tyrosylglycyl-L-tryptophylglycyl-L-
      α-aspartylglycylglycyl-L-tyrosylglycyl-L-seryl-L-α-aspartyl-
      (9CI) (CA INDEX NAME)
OTHER NAMES:
CN   20: PN: JP2004339189 PAGE: 9 unclaimed sequence
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL   14

PATENT ANNOTATIONS (PNTE):
Sequence │Patent
Source   │Reference

```
SEQ     1 GDYGWGDGGY GSDS
          ======== ====
HITS AT:  3-14
MF    C59 H73 N15 O25
SR    CA
LC    STN Files:   CA, CAPLUS, USPATFULL
DT.CA CAplus document type:  Patent
RL.P  Roles from patents:  PRP (Properties)
```
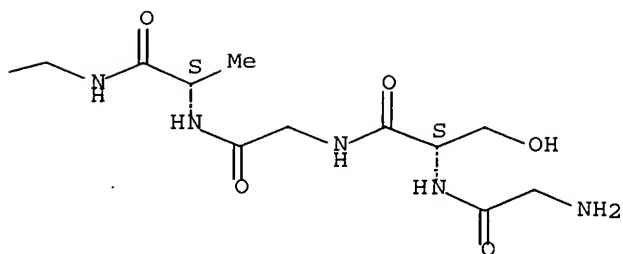
Absolute stereochemistry.
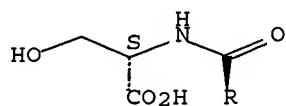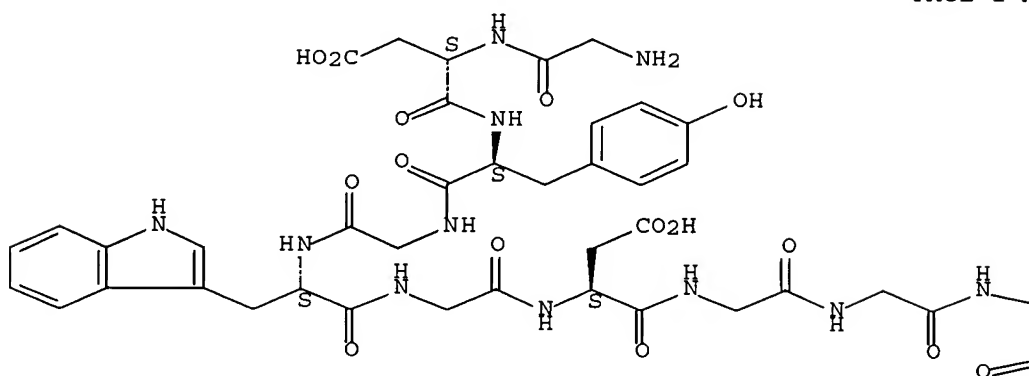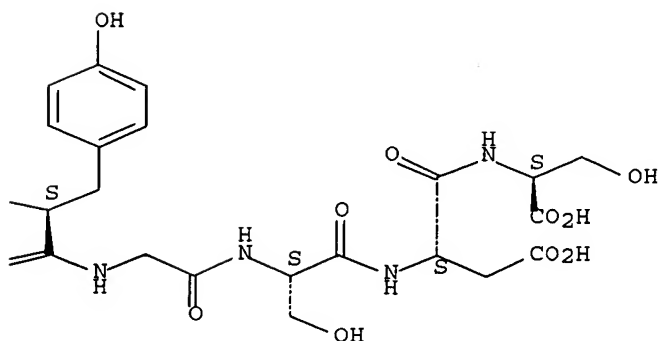
PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

1 REFERENCES IN FILE CA (1907 TO DATE)

L2   ANSWER 10 OF 17 REGISTRY COPYRIGHT 2005 ACS on STN
RN   803731-27-3 REGISTRY
CN   L-Serine, L-serylglycyl-L-alanylglycylglycyl-L-serylglycylglycylglycyl-L-
      tyrosylglycyl-L-tryptophylglycyl-L-α-aspartylglycylglycyl-L-
      tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI) (CA INDEX NAME)
OTHER NAMES:
CN   18: PN: JP2004339189 PAGE: 9 unclaimed sequence
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL   21

PATENT ANNOTATIONS (PNTE):
Sequence |Patent
Source |Reference
=========+=============
Not Given|JP2004339189
         |unclaimed
         |PAGE 9

SEQ   1 SGAGGSGGGY GWGDGGYGSD S
           = ========== =
HITS AT: 10-21
MF   C74 H98 N22 O32
SR   CA
LC   STN Files: CA, CAPLUS, USPATFULL
DT.CA CAplus document type: Patent
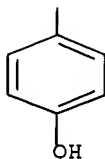RL.P Roles from patents: PRP (Properties)

Absolute stereochemistry.
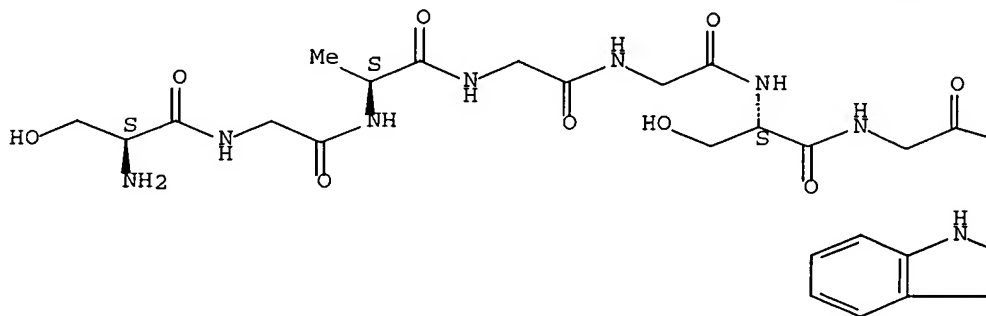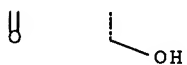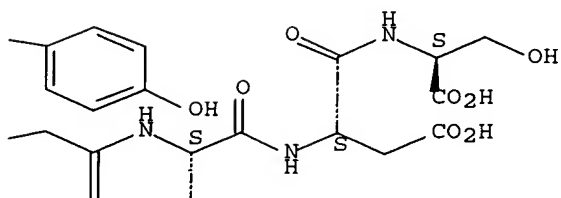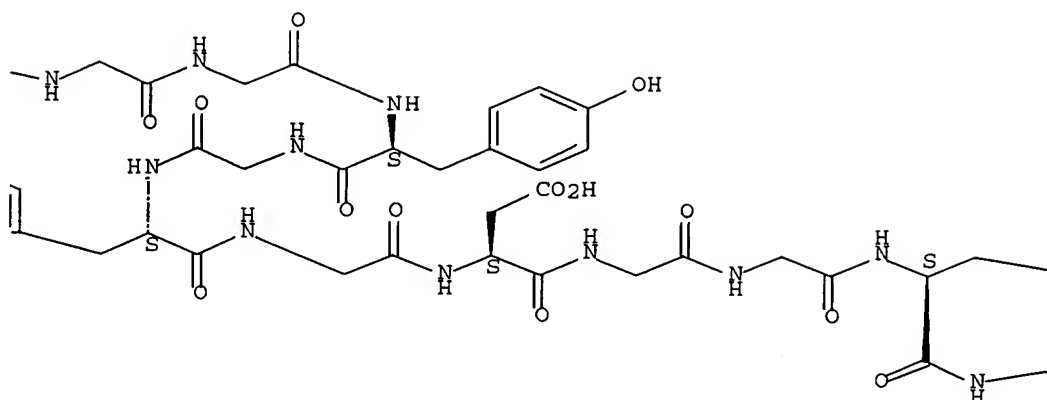
PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

1 REFERENCES IN FILE CA (1907 TO DATE)
1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

L2   ANSWER 11 OF 17   REGISTRY   COPYRIGHT 2005 ACS on STN
RN   799804-74-3   REGISTRY

CN   L-Serine, L-tyrosylglycyl-L-tryptophylglycyl-L-α-
      aspartylglycylglycyl-L-tyrosylglycyl-L-seryl-L-α-aspartyl- (9CI)
      (CA INDEX NAME)
FS   PROTEIN SEQUENCE; STEREOSEARCH
SQL   12

SEQ   1 YGWGDGGYGS DS
            ========== ==
HITS AT: 1-12
MF   C53 H65 N13 O21
SR   CA
LC   STN Files: CA, CAPLUS, USPATFULL
DT.CA CAplus document type: Patent
RL.P Roles from patents: BIOL (Biological study); OCCU (Occurrence); PREP
      (Preparation); USES (Uses)

Absolute stereochemistry.
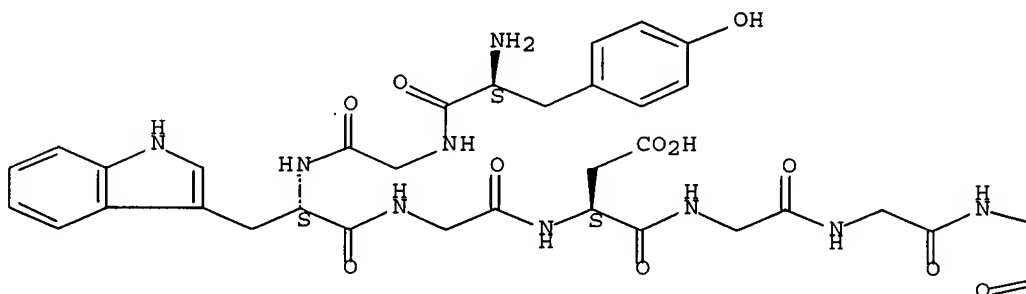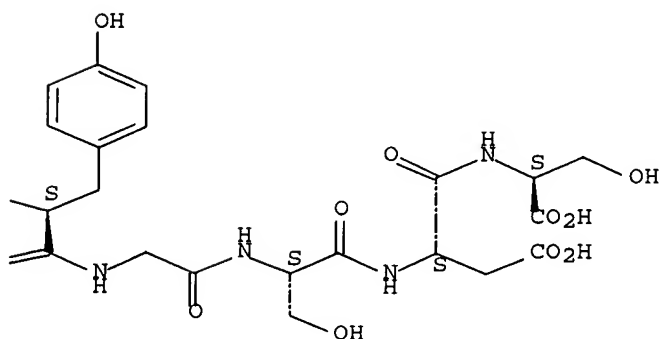
PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

          1 REFERENCES IN FILE CA (1907 TO DATE)
          1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

```
RN   663232-47-1  REGISTRY
CN   Fibroin (Antheraea pernyi strain 741 C-terminal fragment) (9CI)  (CA INDEX
     NAME)
OTHER NAMES:
CN   Fibroin (Chinese Oak Silkworm silk gland strain 741 C-terminal fragment)
FS   PROTEIN SEQUENCE
SQL  405

SEQ      1 RRAGHERAAG SAAAAAAAAA AAASGVGRSG GSYGWGDGGY GSDSAAAAAA
                                         ======== ====
        51 AAAAAAAASG AGGAGVCRGY GGYGSDGSGS AAAAAAAAAA AGSGAGGVGG
       101 GYGWGDGAYG SDSAAAAAAA AAAAAGSGAG GRRGYGAYGS DSSAAAAAAA
       151 AAASGAGGSG GGYGWGDGGY GSDSAAAAAA AAAAAAAAGS GAGGIGGGFG
                                         ======== ====
       201 RGDGGYGSGS SAAAAAAAAA AAARRAGHGR SAGSAAAAAA AAAAAAASGA
       251 GGSGGSYGWD YESYGSGSAA AAAGSGAGGS GGGYGWGDGG YGSGSSAAAA
       301 AAAAAAAGSR RSGHDRAYGA GSAAAAAAAA AAGAGASRQV GIYGTDDGFL
       351 DGGYDSEGSA AAAAAAAAAA ASSSGRSTEG HPLLSICCRP CSHSHSYEAS
       401 RMPVH
HITS AT:    33-44, 163-174
MF   Unspecified
CI   MAN
SR   CA
LC   STN Files:   CA, CAPLUS
DT.CA  CAplus document type:  Journal
RL.NP  Roles from non-patents:  BIOL (Biological study); PRP (Properties)
             1 REFERENCES IN FILE CA (1907 TO DATE)
             1 REFERENCES IN FILE CAPLUS (1907 TO DATE)


L2   ANSWER 13 OF 17  REGISTRY  COPYRIGHT 2005 ACS on STN
RN   481611-80-7  REGISTRY
CN   GenBank AAK51548 (9CI)   (CA INDEX NAME)
OTHER NAMES:
CN   GenBank AAK51548 (Translated from: GenBank AF364332)
FS   PROTEIN SEQUENCE
SQL  436

SEQ      1 RRAGHERAAG SAAAAAAAAA AAASGAGRSG GSYGWGDGGY GSDSAAAAAA
                                         ======== ====
        51 AAAAAAASGS GGSGGYGGYG GYGSDSAAAA AAAAAAAASG AGGAGGYGGY
       101 GGYGSYGSDS AAAAAAAAAA AGSGAGGVGG GYGWGDGGYG SDSAAAAAAA
                                         ======== ===
       151 AAAAAGSGAG GRRGYGAYGS DSSAAAAAAA AAASGAGGSG GGYGWGDGGY
       201 GSDPAAAAAA AAAAAAAAGS GAGGIGGGFG RGDGGYGSGS SAAAAAAAAA
       251 AAARRAGHGR SAGSAAAAAA AAAAAAASGA GGSGGSYGWD YESYGSGSAA
       301 AAAGSGAGGS GGGYGWGDGG YGSGSSAAAA AAAAAAAGSR RSGHDRAYGA
       351 GSAAAAAAAA AAGAGASRQV GIYGTDDGFV LDGGYDSEGS AAAAAAAAAA
       401 AASSSGRSTE GHPLLSICCR PCSHSHSYEA SRISVH
HITS AT:    33-44, 132-143
MF   Unspecified
CI   MAN
SR   GenBank


L2   ANSWER 14 OF 17  REGISTRY  COPYRIGHT 2005 ACS on STN
RN   480176-41-8  REGISTRY
CN   GenBank AAN28165 (9CI)   (CA INDEX NAME)
OTHER NAMES:
CN   GenBank AAN28165 (Translated from: GenBank AY136274)
FS   PROTEIN SEQUENCE
SQL  507
```

```
SEQ       1 MRVIAFVILC CALQYATAKN IHHDEYVDSH GQLVERFTTR KHYERNAATR
         51 PHLSGNERLV ETIVLEEDPY GHEDIYEEDV VIKRVPGASS SAAAASSASA
        101 GSGQTITVER QASHGAGGAA GAAAGAAASS SVRGGGGFYE THDSYSSYGS
        151 DSAAAAAAAA ASGAGGRGHG GYGSDSAAAA AAAAAAAAAA ASGAGGRGHG
        201 GYGSDSAAAA AAAAAAAAAA GSGAGGRGDG GYGWGDGGYG SDSGAAAAAA
                                             ========= ===
        251 AAAAAAASGA GGRGDGGYGR GDGGYGSDSA AAAAAAAAAA AGSGAGGQAT
        301 VVMDGAMAAM VLTRAQQQLA AAAAAASASG AGGSGGSYEW DYGSYGSDSA
        351 AAAAAAAAAA AAGSGAGGVG GGYGRGDGGY GSDSAAAAAA AAAAAAGSGA
        401 GGRGDGGYGW GDGGYGSDSG AAAAAAAAAA AAASGAGGRG DGGYGWGDGG
                 === =========
        451 YGSDPGAAAA AAAAAAAAAS GARGRGDGGY GSGSSAAAAA AAAAAASAAR
        501 RAGHDRA
HITS AT:   232-243, 408-419
MF    Unspecified
CI    MAN
SR    GenBank


L2    ANSWER 15 OF 17  REGISTRY  COPYRIGHT 2005 ACS on STN
RN    404318-03-2  REGISTRY
CN    Fibroin (Antheraea yamamai) (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN    GenBank AAK83145
CN    GenBank AAK83145 (Translated from: GenBank AF325500)
FS    PROTEIN SEQUENCE
SQL   2655


SEQ       1 MRVTAFVILC CALQYATANN LHHHDEYVDN HGQLVERFTT RKHYERNAAT
         51 RPHLSGNERL VETIVLEEDP YGHEDIYEED VVINRVPGAS SSAAAASSAS
        101 AGSGQTIIVE RQASHGAGGA AGAAAGAAAG SSARGGSGFY ETHDSYSSYG
        151 SGSSSAAAAS SGAGGAGGGY GWGDGGYGSD SAAAAAAAAA AAAGSGAGGR
                                = ========== =
        201 GDGGYGSGSS AAAAAAAAAA AARRAGHDHA AGSSGGGYSW DYSSYGSESA
        251 AAAAAAAAAG SGAGGVGGGY GGGDGGYGSG SSAAAAAAAA AAAAARRAGHD
        301 RAAGSAAAAA AAAAAAAASG AGGSGGGYGW GDGGYGSDSA AAAAAAAAAA
                                            === =========
        351 AAGSGAGRAG GDYGWGDGGY GSDSAAAAAA AAAAASGAGG SGGGYGWGDG
                        ======== ====                      ======
        401 GYGSDSAAAA AAAAAASGA GGSGGGYGWG DGGYGSDSAA AAAAAAAAAA
             ======                           ==== ========
        451 GSGAGGRGDG GYGSGSSAAA AAAAAAAAAA RQAGHERAAG SAAAAAAAAA
        501 AAAASGAGGS GRGYGWGDGG YGSDSAAAAA AAAAAAAAGS GAGGAGGDYG
                        ======= =====                      ==
        551 WGDGGYGSDS AAAAAAAAA ASGAGGSGGG YGWGDGGYGS DSAAAAAAAA
             ==========                      ========== ==
        601 AAAAGSGAGG RGDGGYGSGS SAAAAAAAAA AAAARRAGHDR AAGSAAAAAA
        651 AAAAAASGAG GSGGGYGWGD GGYGSDSAAA AAAAAAAAAA GSGAGGAGGD
                        ===== =======
        701 YGWGDGGYGS DSAAAAAAAA AAASGAGGAG GGYGWGDGGY GSDSAAAAAA
             ========== ==                      ======== ====
        751 AAAAAGSGA GGRGDGGYGS GSSAAAAAAA AAAAARRAGH DRAAGSAAAA
        801 AAAAAAAAS GAGGSGGGYG WGDGGYGSDS AAAAAAAAAA AAAGSGAGGA
                        == ==========
        851 GGDYGWGDGG YGSDSAAAAA AAAAAASGAG GSGGGYGWGD GGYGSDSAAA
             ======= =====                      ===== =======
        901 AAAAAAAAAG SGAGGRGDGG YGSGSSAAAA AAAAAAAARR AGHDRAAGSA
        951 AAAAAAAAAA AASGAGGSGG GYGWGDGGYG SDSAAAAAAA AAAAAAGSGA
                                            ========= ===
       1001 GGAGGDYGWG DGGYGSDSAA AAAAAAAAS GAGGAGGGYG WGDGGYGSDS
```

```
                ==== ========                        == ==========
      1051 AAAAAAAAAA AAGSGAGGRG DGGYGSGSSA AAAAAAAAAA ARRAGHDRAA
      1101 GSAAAAAAAA AAAAASGAGG SGGGYGWGDG GYGSDSAAAA AAAAAAAAAG
                                          ====== ======
      1151 SGAGGVGGGY GWGDGGYGSD SAAAAAAAAA AAAASGAGGA GGYGGYGSDS
                =  ========== =
      1201 AAAAAAAAAA AAASSGAGGA GGGYGWGDGG YGSDSAAAAA AAAAAAAGSG
                                          ======= =====
      1251 AGGRGDGGYG SGSSAAAAAA AAAAAAASGA GGSGGGYGWG DGGYGSGSAA
      1301 AAAAAAAAAA AGSGAGGVGG GYGWGDGGYG SDSAAAAAAA AAAAAASGAG
                                          ========= ===
      1351 GAGGYGGYGS DSAAAAAAAA AAAAAGSGAG GAGGGYGWGD GGYGSDSAAA
                                                ===== =======
      1401 AAAAAAAAAA SGAGGRGDGG YGSGSSAAAA AAAAAAAARR AGYDRAAGSA
      1451 AAAAAAAAAA AASGAGGSGG GYGWGDGGYG SDSAAAAAAA AAAAAASGAG
                                          ========= ===
      1501 GAGGYGGYGS DSAAAAAAAA AAAAAGSGAG GAGGGYGWGD GGYGSDSAAA
                                                ===== =======
      1551 AAAAAAAAAA SGAGGRGDGG YGSGSSAAAA AAAAAAAARR AGYDRAAGSA
      1601 AAAAAAAAAA AASGAGGSGG GYGWGDGGYG SDSAAAAAAA AAAAAASGAG
                                          ========= ===
      1651 GAGGYGGYGS DSAAAAAAAA AAAAAGSGAG GAGGGYGWGD GGYGSDSAAA
                                                ===== =======
      1701 AAAAAAAAAG SGAGGRGDGG YGSGSSAAAA AAAAAAAAAR RAGHDRAAGC
      1751 AAAAAAAAAA AASGAGGSGG GYGWGDGGYG SDSAAAAAAA AAAAAGSGA
                                          ========= ===
      1801 GGAGGGYGWG DGGYGSDSAA AAAAAAAAS GAGGTGGGYG WGDGGYGSDS
                ==== ========                        == ==========
      1851 AAAAAAAAAA ASGAGGAGGG YGWGDGGYGS DSAAAAAAAA AAAAGSGAGG
                                ========== ==
      1901 RGDGGYGSGS SAAAAAAAAA AAAARRAGHDR AAGSAAAAAA AAAAAAASGA
      1951 GGSGGGYGWG DGGYGSNSAA AAAAAAAAAA AGSGAGGVGG GYGWGDGGYG
                                                        =========
      2001 SDSAAAAAAA AAAAAAGSGA GGAGGGYGWG DGGYGSDSAA AAAAAAAAAA
                ===                        ==== ========
      2051 GSGAGGRGDG GYGSGSSAAA AAAAAAAAR RAGHDRAAGS AAAAAAAAAA
      2101 AAASGAGRSG GGYGWGDGGY SSDSAAAAAA AAAAAAAGSG AGGVGGGYGW
                                                              ===
      2151 GDGGYGSDSA AAAAAAAAAA SGAGGSGGYG GYGSDSAAAA AAAAAAAAAG
                =========
      2201 SGAGGVGGGY GWGDGGYGSD SAAAAAAAAA AASGAGGSGG YGGYGSDSAA
                = ========== =
      2251 AAAAAAAAAA AGSGAGGVGG GYGWGDGGYG GYGSDSAAAA AAAAAAAGSG
      2301 AGGVGGGYGR GDSGYGSGSS AAAAAAAAAA ARRAGHGRSS GSAAAAAAAA
      2351 AAAASGAGG SGGGYGWDYG SYGSDSAAAA AAAAAAAASS GAGGSGGGYG
      2401 WDYGGYGSDS AAAAAAAAAA AAGSGAGGSG GGYGWGDGGY GSDSAAAAAA
                                          ======== ====
      2451 AAAAAGSGA GGRGDGGYGS GSSAAAAAAA AAAAARRAGH DHAAGSSGGG
      2501 YSWDYSSYGS ESAAAAAAAA AAGSGAGGVG GGYGGGDGGY GSGSSAAAAA
      2551 AAAAAAASRR AGHDRAYGAG SAAAAAAAAA AGAGASRPVG IYGTDDGFVL
      2601 DGGYDSEGSA AAAAAAAAAA ASSSGRSTEG HPLLSICCRP CSHRHSYEAS
      2651 RISVH
HITS AT:    170-181, 328-339, 363-374, 395-406, 427-438, 514-525,
            549-560, 581-592, 666-677, 701-712, 733-744, 819-830,
            854-865, 886-897, 972-983, 1007-1018, 1039-1050,
            1125-1136, 1160-1171, 1224-1235, 1322-1333, 1386-1397,
            1472-1483, 1536-1547, 1622-1633, 1686-1697, 1772-1783,
            1807-1818, 1839-1850, 1871-1882, 1992-2003, 2027-2038,
            2148-2159, 2210-2221, 2433-2444
MF    Unspecified
```

```
CI    MAN
SR    CA
LC    STN Files:   CA, CAPLUS
DT.CA CAplus document type:  Journal
RL.NP Roles from non-patents:  BIOL (Biological study); PRP (Properties)
              1 REFERENCES IN FILE CA (1907 TO DATE)
              1 REFERENCES IN FILE CAPLUS (1907 TO DATE)


L2    ANSWER 16 OF 17  REGISTRY  COPYRIGHT 2005 ACS on STN
RN    336885-96-2  REGISTRY
CN    Fibroin (Antheraea pernyi clone AP2 ) (9CI)   (CA INDEX NAME)
OTHER NAMES:
CN    GenBank AAC32606
CN    GenBank AAC32606 (Translated from: GenBank AF083334)
FS    PROTEIN SEQUENCE
SQL   2639


SEQ      1 MRVIAFVILC CALQYATAKN LRHHDEYVDN HGQLVERFTT RKHFERNAAT
        51 RPHLSGNERL VETIVLEEDP YGHEDIYEED VVIKRVPGAS SSAAAASSAS
       101 AGSGQTIIVE RQASHGAGGA AGAAAGAAAG SSARRGGGFY ETHNSYSSYG
       151 SGSSSAAAGS GAGGVGGGYG SDSAAAAAAA AAAASGAGGS GGYGGYGSDS
       201 AAAAAAAAAA AAAGSGAGGS GGYGGYGGYG SDSAAAAAAA AAAAAAGSSA
       251 GGAGGGYGWG DGGYGSDSAA AAAAAAAAA AGSGAGGSGG YGGYGSDSAA
           ==== ========
       301 AAAAAAAAAA AGSSAGGAGG GYGWGDGGYG SDSAAAAAAA AAAAASSGAG
                      ========= ===
       351 GRGDGGYGSG GSSAAAAAAA AAAAARRAGH DRAAGSAAAA AAAAAAAAAS
       401 GAGGSGGGYG WGDGGYGSDS AAAAAAAAAA AAAGSGAGGA GGGYGWGDSG
              == ==========
       451 YGSDSAAAAA AAAAAAAASG AGGSGGYGGY GSDSAAAAAA AAAAAAAGAG
       501 AGGAGGSYGW GDGGYGSDSA AAAAAAAAAA AGSGAGGRGD GGYGSGSSAA
              === =========
       551 AAAAAAAASA ARRAGHDSAA GSAAAAAAAA AAAAASGAGG SGGGYGWGDG
                                                             ======
       601 GYGSDSAAAA AAAAAAAAAG SGAGGAGGGY GWGDGGYGSD SAAAAAAAAA
           ======                    = ========== =
       651 AAAASGARGS GGYGGYGSDS AAAAAAAAAA AAAGSGAGGV GGGYGWGDGG
                                                            ======
       701 YGSDSAAAAA AAAAAAAGSG AGGRGDGGYG SGSSAAAAAA AAAASAARRA
           =====
       751 GHDSAAGSAA AAAAAAAAAA ASGAGGSGGG YGWGDGGYGS DSAAAAAAAA
                                           ========== ==
       801 AAAAAGSGAG GAGGGYGWGD GGYGSDSAAA AAAAAAAAAA SGARGSGGYG
                      ===== =======
       851 GYGSDSAAAA AAAAAAAAAG SGAGGVGGGY GWGDGGYGSD SAAAAAAAAA
                                 = ========== =
       901 AAAGSGAGGR GDGGYGSGSS AAAAAAAAAA SAARRAGHDS AAGSAAAAAA
       951 AAAAAASGA GGSGGGYGWG DGGYGSDSAA AAAAAAAAAA AGSGAGGAGG
                     ==== ========
      1001 GYGWGDGGYG SDSAAAAAAA AAAAAASGAR GSGGYGGYGS DSAAAAAAAA
           ========= ===
      1051 AAAAAGSGAG GVGGGYGWGD GGYGSDSAAA AAAAAAAAAG SGAGGRGDGG
                      ===== =======
      1101 YGSGSSAAAA AAAAAAAARR AGHDRAAGSA AAAAAAAAAA AASGAGGSGG
      1151 GYGWGDGGYG SDSAAAAAAA AAAAAAASGA GGSGGYGGYG SDSAAAAAAA
           ========= ===
      1201 AAAAAGSGA GGAGGGYGWG DGGYGSDSAA AAAAAAAAAA ASGAGGSGGY
                      ==== ========
      1251 GGYGGYGSDS AAAAAAAAAA AAAGSGAGGA GGGYGWGDGG YGSDSAAAAA
                                                     ======= =====
```

```
      1301 AAAAAAAGSG AGGRGDGGYG SGSSAAAAAA AAAAAAARRA GHDRAAGSAA
      1351 AAAAAAAAAA ASGAGGSGGG YGWGDGGYGS DSAAAAAAAA AAAAAASGAG
                               ========== ==
      1401 GSGGYGGYGS DSAAAAAAAA AAAAAGSGAG GVGGGYGWGD GGYGSDSAAA
                                          =====  =======
      1451 AAAAAAAAAA SGAGGSGGYG GYGSDSAAAA AAAAAAAAAA SGAGGAGGYG
      1501 GYGSDSAAAA AAAAAAAAAG SGAGGAGGGY GWGDGGYGSY SAAAAAAAAA
      1551 AAAGSGAGGR GDGGYGSGSS AAAAAAAAAA AARRAGHDRA AGSAAAAAAA
      1601 AAAAAASGAG GAGGGYGWGD GGYSSDSAAA AAAAAAAAAA GSGAGGAGGG
      1651 YGWGDDGYGS DSAAAAAAAA AAAAGSGAGG RGGGYGWGDG GYGSDSAAAA
                               ======  ======
      1701 AAAAAAAGS GAGGRGDGGY GSGSSAAAAA AAAAAAARRA GHDRAAGSAA
      1751 AAAAAAAAAA SGAGGSGGSY GWGDGGYGSD SAAAAAAAAA AAAASGAGGS
                 = ========== =
      1801 GGYGGYGGYG GYGSDSAAAA AAAAAAAAAA AGSGAGGVGG GYGWGDGGYG
                                                    =========
      1851 SDSAAAAAAA AAAAAGSGAG GRGDGGYGSG SSAAAAAAAA AAAAARRAGHD
           ===
      1901 RAAGSAAAAA AAAAAAAASG AGGAGGGYGW GDGGYGSDSA AAAAAAAAAA
                                          ===  =========
      1951 AAGSGAGGAG GGYGWGDDGY GSDSAAAAAA AAAAAAGSGA GGRGGGYGWG
                                                              ====
      2001 DGGYGSDSAA AAAAAAAAAA GSGAGGRGDG GYGSGSSAAA AAAAAAAAAA
           ========
      2051 RRAGHDRAAG SAAAAAAAAA AAAASGAGGS GGYGGYGGYG SDSAAAAAAA
      2101 AAAAAAGSGA GGAGGYGGYG GYGSDSAAAA AAAAAAAAAG SGAGGVGGGY
                                                              =
      2151 GWGDGGYGSD SAAAAAAAAA AAAAGSGAGG RGDGGYGSGS SAAAAAAAAA
           ========== =
      2201 AAARRAGHER AAGSAAAAAA AAAAAASGAG RSGGSYGWGD GGYGSDSAAA
                                          =====  =======
      2251 AAAAAAAAAA SGAGGSGGYG GYGGYGSDSA AAAAAAAAAA ASGAGGAGGY
      2301 GGYGGYGSYG SDSAAAAAAA AAAAGSGAGG VGGGYGWGDG GYGSDSAAAA
                                          ======  ======
      2351 AAAAAAAGS GAGGRRGYGA YGSDSSAAAA AAAAAASGAG GSGGGYGWGD
                                                              =====
      2401 GGYGSDSAAA AAAAAAAAAA AGSGAGGIGG GFGRGDGGYG SGSSAAAAAA
           =======
      2451 AAAAAARRAG HGRSAGSAAA AAAAAAAAAA SGAGGSGGSY GWDYESYGSG
      2501 SAAAAAGSGA GGSGGGYGWG DGGYGSGSSA AAAAAAAAAA GSRRSGHDRA
      2551 YGAGSAAAAA AAAAAGAGAS RQVGIYGTDD GFVLDGGYDS EGSAAAAAAA
      2601 AAAAASSSGR STEGHPLLSI CCRPCSHSHS YEASRISVH
HITS AT:   257-268, 322-333, 409-420, 508-519, 595-606, 630-641,
           694-705, 781-792, 816-827, 880-891, 967-978, 1002-1013,
           1066-1077, 1152-1163, 1217-1228, 1284-1295, 1371-1382,
           1436-1447, 1685-1696, 1770-1781, 1842-1853, 1928-1939,
           1997-2008, 2150-2161, 2236-2247, 2335-2346, 2396-2407
MF   Unspecified
CI   MAN
SR   CA
LC   STN Files:   CA, CAPLUS
DT.CA  CAplus document type:  Journal
RL.NP  Roles from non-patents:  BIOL (Biological study); PRP (Properties)
             1 REFERENCES IN FILE CA (1907 TO DATE)
             1 REFERENCES IN FILE CAPLUS (1907 TO DATE)


L2   ANSWER 17 OF 17  REGISTRY  COPYRIGHT 2005 ACS on STN
RN   185261-70-5  REGISTRY
CN   Fibroin (Antheraea pernyi clone pApF1.4fibroin C-terminal fragment) (9CI)
     (CA INDEX NAME)
```

```
OTHER NAMES:
CN    GenBank BAA11860
CN    GenBank BAA11860 (Translated from: GenBank D83241)
FS    PROTEIN SEQUENCE
SQL   421


SEQ       1 SGSSAAAAAA AAAASRRAGH ERAAGSAAAA AAAAAAAASG VGRSGGSYGW
                                                                ===
          51 GDGGYGSDSA AAAAAAAAAA AAASGAGGAG VCRGYGGYGS DGSGSAAAAA
             =========
         101 AAAAAAGSGA GGVGGGYGWG DGAYGSDSAA AAAAAAAAAA GSGAGGRRGY
         151 GAYGSDSSAA AAAAAAAASG AGGSGGGYGW GDGGYGSDSA AAAAAAAAAA
                                                       === =========
         201 AAAGSGAGGI GGGFGRGDGG YGSGSSAAAA AAAAAAAARR AGHGRSAGSA
         251 AAAAAAAAAA AASGAGGSGG SYGWDYESYG SGSAAAAAGS GAGGSGGGYG
         301 WGDGGYGSGS SAAAAAAAAA AAGSRRSGHD RAYGAGSAAA AAAAAAAGAG
         351 ASRQVGIYGT DDGFILDGGY DSEGSAAAAA AAAAAAASSS GRSTEGHPLL
         401 SICCRPCSHS HSYEASRMPV H
HITS AT:   48-59, 178-189
MF    Unspecified
CI    MAN
SR    CA
LC    STN Files:   CA, CAPLUS
DT.CA CAplus document type:  Journal
RL.NP Roles from non-patents:  PRP (Properties)
              1 REFERENCES IN FILE CA (1907 TO DATE)
              1 REFERENCES IN FILE CAPLUS (1907 TO DATE)


=> fil caplus
FILE 'CAPLUS' ENTERED AT 08:49:49 ON 12 DEC 2005
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2005 AMERICAN CHEMICAL SOCIETY (ACS)
```

```
FILE COVERS 1907 - 12 Dec 2005  VOL 143 ISS 25
FILE LAST UPDATED: 11 Dec 2005   (20051211/ED)

Effective October 17, 2005, revised CAS Information Use Policies apply.
They are available for your review at:

http://www.cas.org/infopolicy.html
'OBI' IS DEFAULT SEARCH FIELD FOR 'CAPLUS' FILE

=> => d que 13
L1           6 SEA FILE=REGISTRY ABB=ON  PLU=ON  VITTDSDGNE/SQSP
L3           4 SEA FILE=CAPLUS ABB=ON  PLU=ON  L1

=> d que 14
```

=> d 13 .ca 1-4; d .ca 14 1-5

L3   ANSWER 1 OF 4   CAPLUS   COPYRIGHT 2005 ACS on STN
ACCESSION NUMBER: 2004:1035641 CAPLUS Full-text
DOCUMENT NUMBER: 142:33017
TITLE: Cell growth-promoting peptides from silk proteins
INVENTOR(S):   Tsubouchi, Kozo; Yamada, Hiroo
PATENT ASSIGNEE(S):   National Institute of Agrobiological Resources NIAR,
                           Japan
SOURCE: Jpn. Kokai Tokkyo Koho, 27 pp.
                           CODEN: JKXXAF
DOCUMENT TYPE: Patent
LANGUAGE: Japanese
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
JP 2004339189	A2	20041202	JP 2003-406608	20031204
US 2005143296	A1	20050630	US 2004-789494	20040227
CN 1535723	A	20041013	CN 2004-10035241	20040301

PRIORITY APPLN. INFO.:   JP 2003-55048   A 20030228
ED   Entered STN: 03 Dec 2004
AB   Disclosed are cell growth-promoting peptides which comprise 4-40 amino acids
      from noncryst. peptide chains of the silk proteins. The peptides are obtained
      by hydrolyzing silk worm proteins or Antheraea cocoon fibroins and separating
      them by mol. weight fraction. The peptides are effective as cell growth
      promoters, cell adhesives, wound healing promoters, and cell culture matrixes.
      Also claimed is a cosmetic containing the peptides.
IC   ICM C07K014-435
      ICS A61K007-00; A61K038-00; A61K038-17; A61P017-02; C07K001-12;
           C12N005-06; C12P021-06
CC   1-12 (Pharmacology)
      Section cross-reference(s): 62, 63
IT   714954-20-8P   714954-21-9P   799804-72-1P   799804-73-2P
      799804-74-3P   799804-75-4P   799804-76-5P   799804-77-6P
      RL: COS (Cosmetic use); NPO (Natural product occurrence); PAC
      (Pharmacological activity); PNU (Preparation, unclassified); BIOL
      (Biological study); OCCU (Occurrence); PREP (Preparation); USES (Uses)
         (cell growth-promoting peptides from silk proteins)
IT   803731-18-2   803731-19-3   803731-20-6   803731-21-7   803731-22-8
      803731-23-9   803731-24-0   803731-25-1   803731-26-2   803731-27-3
      803731-28-4   803731-29-5   803731-30-8   803731-31-9   803731-32-0
      803731-33-1   803731-34-2   803731-36-4   803731-37-5   803731-38-6
      803731-39-7   803731-40-0   803731-41-1   803731-44-4   803731-46-6
      803731-47-7   803731-48-8   803731-49-9   803731-50-2   803731-51-3
      803731-52-4   803731-53-5   803731-54-6   803731-55-7   803731-56-8
      803823-75-8   803823-76-9   803823-77-0   803823-78-1
      803823-79-2   803823-80-5   803823-81-6   803823-82-7   803823-83-8
      RL: PRP (Properties)
         (unclaimed sequence; cell growth-promoting peptides from silk proteins)

L3   ANSWER 2 OF 4   CAPLUS   COPYRIGHT 2005 ACS on STN
ACCESSION NUMBER: 2003:838448 CAPLUS Full-text
DOCUMENT NUMBER: 141:82207
TITLE: Identification of fibroin-derived peptides enhancing

AB   The authors previously reported that the fibroin of the silkworm Bombyx mori
      enhanced the proliferation of cultured human skin fibroblasts. In this work,
      the fibroin was digested by chymotrypsin, and the resulting peptide fragments
      were fractionated and assayed for their biol. activity. Two peptides that
      promoted fibroblast growth were isolated and identified to be VITTDSDGNE and
      NINDFDED. Both sequences are found in the N-terminal region of the fibroin
      polypeptide and are thought to be the active principle of fibroblast growth-
      promoting activity.
CC   1-12 (Pharmacology)
      Section cross-reference(s): 12
IT   714954-20-8   714954-21-9
      RL: PAC (Pharmacological activity); THU (Therapeutic use); BIOL
      (Biological study); USES (Uses)
          (fibroin-derived peptides enhancing proliferation of cultured human
          skin fibroblasts)
REFERENCE COUNT: 12   THERE ARE 12 CITED REFERENCES AVAILABLE FOR THIS
                                    RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L3   ANSWER 3 OF 4   CAPLUS   COPYRIGHT 2005 ACS on STN
AB   We sequenced an 80 kb DNA region containing the complete sequence of the
      silkworm Bombyx mori fibroin gene and its flanking, especially the upstream,
      regions (.apprx.62 kb). About 30% of the 62 kb upstream region is composed of
      repetitive elements including short interspersed elements Bm1, long
      interspersed elements L1Bm and mariner-like elements Bmmar1 which are
      widespread over the silkworm genome. This 62 kb region is also enriched of
      commonly considered matrix association region (MAR) motifs. A total of 25
      individual MAR recognition signatures (MRSs) were identified, with 24 at the
      upstream and one at the downstream region. Combining two newly developed MAR
      prediction programs (MAR-finder and Chrclass), ten candidate MARs were
      predicted, with five containing MRS and seven related to the repetitive
      elements. The wide distribution of nested repetitive elements, candidate
      MARs, DNase I hypersensitive sites and other potential regulatory factors

recognition sites indicates this region is probably a unique huge cis-acting
element contributing to the regulation of the spatial and temporal specificity
and efficiency of fibroin gene expression.
CC   3-3 (Biochemical Genetics)
      Section cross-reference(s): 6, 12
IT   303229-60-9, Fibroin heavy chain (silkworm strain p50)
      RL: BSU (Biological study, unclassified); PRP (Properties); BIOL
      (Biological study)
          (amino acid sequence; 62-kb upstream region of Bombyx mori fibroin
      heavy chain gene has clustered repetitive elements and candidate matrix
      association regions)
REFERENCE COUNT: 33   THERE ARE 33 CITED REFERENCES AVAILABLE FOR THIS
                                 RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L3   ANSWER 4 OF 4   CAPLUS   COPYRIGHT 2005 ACS on STN
ACCESSION NUMBER: 2000:472155   CAPLUS Full-text
DOCUMENT NUMBER: 133:330213
TITLE: Fine organization of Bombyx mori fibroin heavy chain
                           gene
AUTHOR(S):   Zhou, Cong-Zhao; Confalonieri, Fabrice; Medina,
                           Nadine; Zivanovic, Yvan; Esnault, Catherine; Yang,
                           Tie; Jacquet, Michel; Janin, Joel; Duguet, Michel;
                           Perasso, Roland; Li, Zhen-Gang
CORPORATE SOURCE: Institut de Genetique et Microbiologie and Laboratoire
                           de Biologie Cellulaire 4, Universite Paris-Sud et
                           CNRS, Orsay, 91405, Fr.
SOURCE: Nucleic Acids Research (2000), 28(12), 2413-2419
                           CODEN: NARHAD; ISSN: 0305-1048
PUBLISHER: Oxford University Press
DOCUMENT TYPE: Journal
LANGUAGE: English
ED   Entered STN: 13 Jul 2000
AB   The complete sequence of the Bombyx mori fibroin gene has been determined by
      means of combining a shotgun sequencing strategy with phys. map-based
      sequencing procedures. It consists of two exons (67 and 15 750 bp, resp.) and
      one intron (971 bp). The fibroin coding sequence presents a spectacular
      organization, with a highly repetitive and G-rich (.apprx.45%) core flanked by
      non-repetitive 5' and 3' ends. This repetitive core is composed of alternate
      arrays of 12 repetitive and 11 amorphous domains. The sequences of the
      amorphous domains are evolutionarily conserved and the repetitive domains
      differ from each other in length by a variety of tandem repeats of subdomains
      of .apprx.208 bp which are reminiscent of the repetitive nucleosome
      organization. A typical composition of a subdomain is a cluster of repetitive
      units, Ua, followed by a cluster of units, Ub, (with a Ua:Ub ratio of 2:1)
      flanked by conserved boundary elements at the 3' end. Moreover some repeats
      are also perfectly conserved at the peptide level indicating that the
      evolutionary pressure is not identical along the sequence. A tentative model
      for the constitution and evolution of this unusual gene is discussed.
CC   3-3 (Biochemical Genetics)
      Section cross-reference(s): 12
IT   303229-60-9
      RL: BSU (Biological study, unclassified); PRP (Properties); BIOL
      (Biological study)
          (amino acid sequence; fine organization of Bombyx mori fibroin heavy
      chain gene)
REFERENCE COUNT: 29   THERE ARE 29 CITED REFERENCES AVAILABLE FOR THIS
                                 RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

ACCESSION NUMBER: 2004:1035641 CAPLUS Full-text
DOCUMENT NUMBER: 142:33017
TITLE: Cell growth-promoting peptides from silk proteins
INVENTOR(S):   Tsubouchi, Kozo; Yamada, Hiroo
PATENT ASSIGNEE(S):   National Institute of Agrobiological Resources NIAR,
                        Japan
SOURCE: Jpn. Kokai. Tokkyo Koho, 27 pp.
                        CODEN: JKXXAF
DOCUMENT TYPE: Patent
LANGUAGE: Japanese
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
JP 2004339189	A2	20041202	JP 2003-406608	20031204
US 2005143296	A1	20050630	US 2004-789494	20040227
CN 1535723	A	20041013	CN 2004-10035241	20040301

PRIORITY APPLN. INFO.:   JP 2003-55048   A 20030228
ED   Entered STN: 03 Dec 2004
AB   Disclosed are cell growth-promoting peptides which comprise 4-40 amino acids
     from noncryst. peptide chains of the silk proteins. The peptides are obtained
     by hydrolyzing silk worm proteins or Antheraea cocoon fibroins and separating
     them by mol. weight fraction. The peptides are effective as cell growth
     promoters, cell adhesives, wound healing promoters, and cell culture matrixes.
     Also claimed is a cosmetic containing the peptides.
IC   ICM C07K014-435
     ICS A61K007-00; A61K038-00; A61K038-17; A61P017-02; C07K001-12;
          C12N005-06; C12P021-06
CC   1-12 (Pharmacology)
     Section cross-reference(s): 62, 63
IT   714954-20-8P   714954-21-9P   799804-72-1P   799804-73-2P
     799804-74-3P   799804-75-4P   799804-76-5P   799804-77-6P
     RL: COS (Cosmetic use); NPO (Natural product occurrence); PAC
     (Pharmacological activity); PNU (Preparation, unclassified); BIOL
     (Biological study); OCCU (Occurrence); PREP (Preparation); USES (Uses)
        (cell growth-promoting peptides from silk proteins)
IT   803731-18-2   803731-19-3   803731-20-6   803731-21-7   803731-22-8
     803731-23-9   803731-24-0   803731-25-1   803731-26-2 803731-27-3
     803731-28-4 803731-29-5   803731-30-8 803731-31-9
     803731-32-0 803731-33-1 803731-34-2   803731-36-4
     803731-37-5 803731-38-6   803731-39-7 803731-40-0
     803731-41-1 803731-44-4   803731-46-6   803731-47-7
     803731-48-8   803731-49-9   803731-50-2   803731-51-3   803731-52-4
     803731-53-5 803731-54-6   803731-55-7   803731-56-8
     803823-75-8   803823-76-9   803823-77-0 803823-78-1
     803823-79-2   803823-80-5   803823-81-6   803823-82-7   803823-83-8
     RL: PRP (Properties)
      (unclaimed sequence; cell growth-promoting peptides from silk proteins)

ACCESSION NUMBER: 2003:997528 CAPLUS Full-text
DOCUMENT NUMBER: 140:195109
TITLE: Variation and characterization analysis of partial
                        fragment of fibroin gene from silkworm, Antheraea
                        pernyi
AUTHOR(S):   Li, Wenli; Jin, Liji; An, Lija
CORPORATE SOURCE: Department of Bioengineering Chemistry, Dalian

AB   A 1.4Kb DNA fragment containing 3' flanking sequence of fibroin gene of
      silkworm, Antheraea pernyi, was obtained from the silk gland's mRNA of 5th
      larva. Anal. of this sequence with another A. pernyi fibroin protein
      (accession Number D83241) revealed that it consists of a completely open
      reading frame (ORF), which includes 14 polyalanine-containing units (motifs)
      and 100bp 3'-UTR. The sequence of the predicted amino acid reveals the
      highest level of overall identity (90%) with D83241. It was found that it
      loses a repeat region at the upstream of TAA codon and some mutations. A
      putative polyadenylation signal AATAAA tail was found in position 1300, which
      follows the termination codon.
CC   6-3 (General Biochemistry)
      Section cross-reference(s): 3, 11
IT   663232-47-1
      RL: BSU (Biological study, unclassified); PRP (Properties); BIOL
      (Biological study)
          (amino acid sequence; partial sequence and conserved protein motifs of
      fibroin from silkworm (Antheraea pernyi))
REFERENCE COUNT: 7   THERE ARE 7 CITED REFERENCES AVAILABLE FOR THIS
                                    RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L4   ANSWER 3 OF 5   CAPLUS   COPYRIGHT 2005 ACS on STN
ACCESSION NUMBER: 2001:669212   CAPLUS Full-text
DOCUMENT NUMBER: 136:242651
TITLE: Cloning of the fibroin gene from the oak silkworm,
                           Antheraea yamamai and its complete sequence
AUTHOR(S):   Hwang, Jae-Sam; Lee, Jin-Sung; Goo, Tae-Won; Yun,
                           Eun-Young; Lee, Kwang-Sik; Kim, Yong-Sung; Jin,
                           Byung-Rae; Lee, Sang-Mong; Kim, Keun-Young; Kang,
                           Seok-Woo; Suh, Dong-Sang
CORPORATE SOURCE: Department of Sericulture and Entomology, National
                           Institute of Agricultural Science and Technology, RDA,
                           Suwon, 441-100, S. Korea
AB   The nucleotide sequences containing an entire genomic region and 5' upstream
      region of Antheraea yamamai fibroin gene have been determined The gene
      consists of an initial exon encoding 14 amino acids, an intron (150 bp), and a
      long second exon coding for 2641 amino acids. The fibroin coding sequence
      shows a specialized organization with a highly repetitive region flanked by
      non repetitive 5' and 3' ends. Northern blot analyses confirmed that fibroin
      gene is actively expressed in the posterior silk gland of the final instar
      larvae of Antheraea yamamai.
CC   3-3 (Biochemical Genetics)
      Section cross-reference(s): 6, 12
IT   404318-03-2, Fibroin (Antheraea yamamai)
      RL: BSU (Biological study, unclassified); PRP (Properties); BIOL
      (Biological study)
          (amino acid sequence; sequence of the fibroin gene from the oak

silkworm, Antheraea yamamai)
REFERENCE COUNT: 12   THERE ARE 12 CITED REFERENCES AVAILABLE FOR THIS
                               RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L4   ANSWER 4 OF 5 CAPLUS   COPYRIGHT 2005 ACS on STN
ACCESSION NUMBER: 2000:831431 CAPLUS Full-text
DOCUMENT NUMBER: 134:362047
TITLE: Dynamic rearrangement within the Antheraea pernyi silk
                         fibroin gene is associated with four types of
                         repetitive units
AUTHOR(S):   Sezutsu, Hideki; Yukuhiro, Kenji
CORPORATE SOURCE: Department of Insect Genetic Breeding, National
                         Institute of Sericultural and Entomological Science,
                         Tsukuba, 305-8634, Japan
SOURCE: Journal of Molecular Evolution (2000), 51(4), 329-338
                         CODEN: JMEVAU; ISSN: 0022-2844
PUBLISHER: Springer-Verlag New York Inc.
DOCUMENT TYPE: Journal
LANGUAGE: English
ED   Entered STN: 29 Nov 2000
AB   We characterized a full-length gene encoding wild silkmoth Antheraea pernyi
      fibroin (Ap-fibroin) to clarify the conformation of repetitive sequences. The
      gene consisted of a first exon encoding 14 amino acid residues, a short intron
      (120 bp), and a long second exon encoding 2,625 amino acid residues. Three
      amino acids, alanine, glycine, and serine, amounted to 81% of the Ap-fibroin
      sequence. The Ap-fibroin, except for 155 residues of the amino terminus, was
      composed of 80 tandemly arranged polyalanine-containing units (motifs). A
      motif was a doublet of a polyalanine block (PAB) and a nonpolyalanine block
      (NPAB). Seventy-eight of the 80 motifs were classified into four types based
      on differences in the NPAB sequences. Although resp. motifs were
      significantly conserved, many rearrangements were observed within the second
      exon, i.e., the triplication of a 558-bp-long sequence and other duplication
      events of shorter sequences. Chi-like sequences, GCTGGAG, might contribute to
      the rearrangement within the gene as described in human minisatellite loci,
      because they were found at specific sites of NPAB-encoding sequences in three
      of four types of motifs. The present results support the idea that the Ap-
      fibroin gene is unstable like minisatellite sequences and that the evolution
      of this gene is strongly associated with its instability.
CC   3-3 (Biochemical Genetics)
      Section cross-reference(s): 6, 12
IT   336885-96-2, Fibroin (Antheraea pernyi clone AP2)
      RL: BSU (Biological study, unclassified); PRP (Properties); BIOL
      (Biological study)
          (amino acid sequence; dynamic rearrangement within the Antheraea pernyi
      silk fibroin gene is associated with four types of repetitive units)
REFERENCE COUNT: 24   THERE ARE 24 CITED REFERENCES AVAILABLE FOR THIS
                               RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L4   ANSWER 5 OF 5 CAPLUS   COPYRIGHT 2005 ACS on STN
ACCESSION NUMBER: 1997:15804 CAPLUS Full-text
DOCUMENT NUMBER: 126:55749
TITLE: Preferential codon usage and two types of repetitive
                         motifs in the fibroin gene of the Chinese oak
                         silkworm, Antheraea pernyi
AUTHOR(S):   Yukuhiro, K.; Kanda, T.; Tamura, T.
CORPORATE SOURCE: Inst. Sericultural Entomological Science, Ministry
                         Agriculture Fisheries and Forestry, Ibaraki, 305,
                         Japan
SOURCE: Insect Molecular Biology (1997), 6(1), 89-95
                         CODEN: IMBIE3; ISSN: 0962-1075

AB   In this paper we describe the peculiar structures and preferential codon usage
      found in wild silkworm fibroin genes. We determined a 1350 bp nucleotide
      sequence from the Chinese oak silkworm, Antheraea pernyi. The deduced amino
      acid sequence was partitioned into thirteen polyalanine-containing repetitive
      motifs, which was one of the characteristic of Antheraea fibroins. Eleven of
      these arrays can be classified into two types of motifs depending on
      difference in amino acid sequences following polyalanine. Repetitive motifs
      structurally similar to those of A. pernyi were detected in a homolog of the
      Japanese oak silkworm, Antheraea yamamai. The most remarkable feature of this
      study was preferential codon usage, especially seen in alanine synonymous
      codons within both homologs of Antheraea: isocodon GCA most frequently
      occurred in alanine isocodons. In contrast, GCU isocodon was the most
      abundant in Bombyx mori fibroin heavy chain that lacks polyalanine arrays.
      This result strongly suggests different modes of selective constraint between
      the two types of fibroin gene. The similar finding that GCA isocodon was most
      frequent in two dragline silk sequences of the spider, Nephila clavipes, is
      consistent with our results because of the repetitive polyalanine-containing
      arrays seen in spider dragline silk.
CC   3-3 (Biochemical Genetics)
      Section cross-reference(s): 6, 12
IT   185261-70-5
      RL: PRP (Properties)
         (amino acid sequence; preferential codon usage and two types of
         repetitive motifs in the fibroin gene of the Chinese oak silkworm,
         Antheraea pernyi)
REFERENCE COUNT: 24   THERE ARE 24 CITED REFERENCES AVAILABLE FOR THIS
                                 RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:04:00 ; Search time 132.727 Seconds
                                              (without alignments)
                                              53.156 Million cell updates/sec

Title: US-10-789-494B-1
Perfect score: 51
Sequence: 1 VITTDSDGNE 10

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   UniProt_05.80:*
                 1: uniprot_sprot:*
                 2: uniprot_trembl:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	51	100.0	178	1	FIBH_BOMMA	Q99050 bombyx mand
2	51	100.0	5263	1	FIBH_BOMMO	P05790 bombyx mori
3	43	84.3	317	2	Q9VIX8_DROME	Q9vix8 drosophila
4	42	82.4	907	2	Q733M3_BACC1	Q733m3 bacillus ce
5	42	·82.4	11103	2	Q54CU4_DICDI	Q54cu4 dictyosteli
6	41	80.4	150	2	Q742N8_MYCPA	Q742n8 mycobacteri
7	41	80.4	361	2	Q8X078_NEUCR	Q8x078 neurospora
8	40	78.4	318	1	CYPR_YEAST	P25334 saccharomyc
9	40	78.4	423	2	Q4IDY7_GIBZE	Q4idy7 gibberella
10	40	78.4	1312	2	Q9U113_LEIMA	Q9u113 leishmania
11	40	78.4	1326	2	Q6FR84_CANGA	Q6fr84 candida gla
12	40	78.4	1597	2	Q6BXP0_DEBHA	Q6bxp0 debaryomyce
13	39	76.5	121	2	Q6Q4F2_ACTAC	Q6q4f2 actinobacil
14	39	76.5	336	2	Q7T6X9_MIMIV	Q7t6x9 mimivirus.
15	39	76.5	359	2	Q84V38_VITVI	Q84v38 vitis vinif

16	39	76.5	467	2	Q53XD3_DROME	Q53xd3	drosophila
17	39	76.5	475	1	IF2G_DROME	Q24208	drosophila
18	39	76.5	559	1	MDL1_PRUDU	O24243	prunus dulc
19	39	76.5	559	2	Q7XJE8_PRUDU	Q7xje8	prunus dulc
20	39	76.5	563	2	Q7PL55_DROME	Q7pl55	drosophila
21	39	76.5	744	2	Q7PL56_DROME	Q7pl56	drosophila
22	39	76.5	907	2	Q4MNH3_BACCE	Q4mnh3	bacillus ce
23	39	76.5	907	2	Q81AF6_BACCR	Q81af6	bacillus ce
24	39	76.5	907	2	Q6HFI4_BACHK	Q6hfi4	bacillus th
25	39	76.5	907	2	Q637L2_BACCZ	Q637l2	bacillus ce
26	39	76.5	907	2	Q81Y80_BACAN	Q81y80	bacillus an
27	39	76.5	1257	2	Q6BVF7_DEBHA	Q6bvf7	debaryomyce
28	39	76.5	1511	2	Q8A0B0_BACTN	Q8a0b0	bacteroides
29	39	76.5	2233	2	Q81890_PI3B	Q81890	bovine para
30	38	74.5	237	1	LECA_DIOGU	P81637	dioclea gui
31	38	74.5	244	2	Q8K8B0_STRP3	Q8k8b0	streptococc
32	38	74.5	244	2	Q9A100_STRPY	Q9a100	streptococc
33	38	74.5	244	2	Q8P215_STRP8	Q8p215	streptococc
34	38	74.5	252	2	Q878E9_STRP3	Q878e9	streptococc
35	38	74.5	252	2	Q5XDA1_STRP6	Q5xda1	streptococc
36	38	74.5	257	2	Q8UFQ2_AGRT5	Q8ufq2	agrobacteri
37	38	74.5	296	2	Q9NF63_CAEEL	Q9nf63	caenorhabdi
38	38	74.5	328	2	Q6GUL4_9BACT	Q6gul4	prevotella
39	38	74.5	367	2	Q9HMM3_HALSA	Q9hmm3	halobacteri
40	38	74.5	373	2	Q6VSY7_9VIRU	Q6vsy7	vibrio para
41	38	74.5	383	2	Q5MK34_9PAST	Q5mk34	pasteurella
42	38	74.5	391	2	Q5MK37_9PAST	Q5mk37	pasteurella
43	38	74.5	392	2	Q5MK35_9PAST	Q5mk35	pasteurella
44	38	74.5	394	2	Q5MK32_9PAST	Q5mk32	pasteurella
45	38	74.5	437	2	Q22993_CAEEL	Q22993	caenorhabdi
46	38	74.5	510	2	Q5B4P9_EMENI	Q5b4p9	aspergillus
47	38	74.5	550	1	PME22_LYCES	Q96575	lycopersico
48	38	74.5	593	2	Q6A5C6_PROAC	Q6a5c6	propionibac
49	38	74.5	683	1	YPR4_CAEEL	Q20059	caenorhabdi
50	38	74.5	782	2	Q93SH4_BRAJA	Q93sh4	bradyrhizob
51	38	74.5	788	2	Q89EK1_BRAJA	Q89ek1	bradyrhizob
52	38	74.5	917	2	Q88UJ0_LACPL	Q88uj0	lactobacill
53	38	74.5	953	1	LKA11_PASHA	P55118	pasteurella
54	38	74.5	953	1	LKA1A_PASHA	P16535	pasteurella
55	38	74.5	953	1	LKA1B_PASHA	Q7bhi8	pasteurella
56	38	74.5	953	1	LKA2D_PASHA	Q9ev29	pasteurella
57	38	74.5	953	1	LKA7A_PASHA	P0c084	pasteurella
58	38	74.5	953	1	LKTA6_PASHA	P0c083	pasteurella
59	38	74.5	953	1	LKTA8_PASHA	Q9ev34	pasteurella
60	38	74.5	953	1	LKTA_MANGL	Q9etx2	mannheimia
61	38	74.5	953	2	Q6TB03_9PAST	Q6tb03	mannheimia
62	38	74.5	1012	1	UBA1_SCHPO	O94609	schizosacch
63	38	74.5	3444	2	Q4Q439_LEIMA	Q4q439	leishmania
64	37	72.5	127	2	Q8C3K4_MOUSE	Q8c3k4	mus musculu
65	37	72.5	147	2	Q6ENJ8_ORYSA	Q6enj8	oryza sativ
66	37	72.5	169	2	Q5FGQ4_EHRRG	Q5fgq4	ehrlichia r
67	37	72.5	177	2	Q6VCX3_EHRRU	Q6vcx3	ehrlichia r
68	37	72.5	177	2	Q5HA06_EHRRW	Q5ha06	ehrlichia r
69	37	72.5	242	2	Q22207_CAEEL	Q22207	caenorhabdi
70	37	72.5	349	2	Q22512_CAEEL	Q22512	caenorhabdi
71	37	72.5	507	1	FRS2_MOUSE	Q8c180	mus musculu
72	37	72.5	509	2	Q8UVU3_XENLA	Q8uvu3	xenopus lae

73	37	72.5	509	2	Q90ZF5_XENLA	Q90zf5 xenopus lae
74	37	72.5	509	2	Q7ZWM2_XENLA	Q7zwm2 xenopus lae
75	37	72.5	511	1	FRS2_HUMAN	Q8wu20 homo sapien
76	37	72.5	580	2	Q4H4Q5_9DEIO	Q4h4q5 deinococcus
77	37	72.5	585	2	Q6QPZ0_9LACT	Q6qpz0 lactococcus
78	37	72.5	604	2	Q50SY3_ENTHI	Q50sy3 entamoeba h
79	37	72.5	814	2	Q648S9_9ARCH	Q648s9 uncultured
80	37	72.5	817	2	Q6ZPN1_MOUSE	Q6zpn1 mus musculu
81	37	72.5	826	2	Q8IY15_HUMAN	Q8iy15 homo sapien
82	37	72.5	862	2	Q8NTA1_CORGL	Q8nta1 corynebacte
83	37	72.5	1019	2	Q7UWL9_RHOBA	Q7uwl9 rhodopirell
84	37	72.5	1727	2	Q68FD9_MOUSE	Q68fd9 mus musculu
85	37	72.5	1865	2	Q9HCM3_HUMAN	Q9hcm3 homo sapien
86	37	72.5	1902	2	Q9AIQ2_LACLC	Q9aiq2 lactococcus
87	37	72.5	2630	2	Q6ALE1_DESPS	Q6ale1 desulfotale
88	37	72.5	4190	2	Q6K796_ORYSA	Q6k796 oryza sativ
89	36	70.6	103	2	Q595F8_MYCGA	Q595f8 mycoplasma
90	36	70.6	103	2	Q595G2_MYCGA	Q595g2 mycoplasma
91	36	70.6	103	2	Q595J5_MYCGA	Q595j5 mycoplasma
92	36	70.6	103	2	Q595J9_MYCGA	Q595j9 mycoplasma
93	36	70.6	103	2	Q595L9_MYCGA	Q595l9 mycoplasma
94	36	70.6	103	2	Q595M4_MYCGA	Q595m4 mycoplasma
95	36	70.6	103	2	Q595M6_MYCGA	Q595m6 mycoplasma
96	36	70.6	110	2	Q4LDH1_MYCGA	Q4ldh1 mycoplasma
97	36	70.6	157	2	Q7NFK5_GLOVI	Q7nfk5 gloeobacter
98	36	70.6	183	2	O10618_9NUCL	O10618 helicoverpa
99	36	70.6	201	2	Q91BY7_9NUCL	Q91by7 helicoverpa
100	36	70.6	203	2	Q77LZ1_9NUCL	Q77lz1 helicoverpa

ALIGNMENTS

RESULT 1
FIBH_BOMMA
ID   FIBH_BOMMA   STANDARD;   PRT;   178 AA.
AC   Q99050;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-MAY-2005 (Rel. 47, Last annotation update)
DE   Fibroin heavy chain precursor (Fib-H) (H-fibroin) (Fragment).
GN   Name=FIBH;
OS   Bombyx mandarina (Wild silk moth) (Wild silkworm).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC   Bombycidae; Bombyx.
OX   NCBI_TaxID=7092;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Posterior silk gland;
RA   Kusuda J., Tazima Y., Onimaru K., Ninaki O., Suzuki Y.;
RT   "The sequence around the 5' end of the fibroin gene from the wild
RT   silkworm, Bombyx mandarina, and comparison with that of the
RT   domesticated species, B. mori.";
RL   Mol. Gen. Genet. 203:359-364(1986).
CC   -!- FUNCTION: Core component of the silk filament; a strong, insoluble
CC   and chemically inert fiber.

```
CC   -!- SUBUNIT: Silk fibroin elementary unit consists in a disulfide-
CC       linked heavy and light chain and a p25 glycoprotein in molar
CC       ratios of 6:6:1. This results in a complex of approximately 2.3
CC       MDa.
CC   -!- TISSUE SPECIFICITY: Produced exclusively in the posterior (PSG)
CC       section of silk glands, which are essentially modified salivary
CC       glands.
CC   -!- DOMAIN: Composed of antiparallel beta sheets. The strands of the
CC       beta sheets run parallel to the fiber axis. Long stretches of silk
CC       fibroin are composed of microcrystalline arrays of (-Gly-Ser-Gly-
CC       Ala-Gly-Ala-)n interrupted by regions containing bulkier residues.
CC       The fiber is composed of microcrystalline arrays alternating with
CC       amorphous regions.
CC   -!- PTM: The interchain disulfide bridge is essential for the
CC       intracellular transport and secretion of fibroin.
CC   --------------------------------------------------------------------
CC   This Swiss-Prot entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use as long as its content is in no way modified and this statement is not
CC   removed.
CC   --------------------------------------------------------------------
DR   EMBL; X03973; CAA27612.1; -; Genomic_DNA.
KW   Repeat; Signal; Silk.
FT   SIGNAL        1     21        Potential.
FT   CHAIN        22    >178       Fibroin heavy chain.
FT   REGION      149    >178       Highly repetitive.
FT   CONFLICT     10     10        C -> V (in Ref. 1; CAA27612).
FT   NON_TER     178    178
SQ   SEQUENCE   178 AA;  18326 MW;  8E15C7E7A9682940 CRC64;

   Query Match               100.0%;  Score 51;  DB 1;  Length 178;
   Best Local Similarity     100.0%;  Pred. No. 0.24;
   Matches    10;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 VITTDSDGNE 10
              ||||||||||
Db         85 VITTDSDGNE 94




Search completed: December  8, 2005, 08:15:35
Job time : 137.727 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:04:51 ; Search time 21.3636 Seconds
                                          (without alignments)
                                          45.038 Million cell updates/sec

Title: US-10-789-494B-1
Perfect score: 51
Sequence: 1 VITTDSDGNE 10

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   PIR_80:*
                1: pir1:*
                2: pir2:*
                3: pir3:*
                4: pir4:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                              SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	40	78.4	318	1	CSBYC3	peptidylprolyl iso
2	39	76.5	475	1	S46941	translation initia
3	38	74.5	237	2	A45587	lectin - Dioclea l
4	38	74.5	257	2	AI2741	conserved hypothet
5	38	74.5	257	2	H97522	hypothetical prote
6	38	74.5	296	2	T31582	hypothetical prote
7	38	74.5	367	2	H84397	hypothetical prote
8	38	74.5	437	2	T29330	hypothetical prote
9	38	74.5	683	2	T21810	hypothetical prote
10	38	74.5	686	2	T21808	hypothetical prote
11	38	74.5	953	1	B30169	leukotoxin A - Pas
12	38	74.5	1011	2	T50344	poly(A)+ RNA trans
13	38	74.5	1012	2	T52000	poly(A)+ RNA trans

14	37	72.5	242	2	T16804	hypothetical prote
15	37	72.5	349	2	T16882	hypothetical prote
16	36	70.6	455	1	A69753	glucarate dehydrat
17	36	70.6	1122	2	T18346	MGC1 protein precu
18	36	70.6	1409	2	S74916	alkaline phosphata
19	36	70.6	1481	2	S28669	pullulanase (EC 3.
20	36	70.6	1816	2	F83901	hypothetical prote
21	35	68.6	161	2	S67178	translation initia
22	35	68.6	311	2	T24947	hypothetical prote
23	35	68.6	392	2	AD2360	hypothetical prote
24	35	68.6	409	2	T25935	hypothetical prote
25	35	68.6	415	2	S37340	flo protein homolo
26	35	68.6	491	2	T30590	alkylhalidase homo
27	35	68.6	546	2	S46527	pectinesterase (EC
28	35	68.6	914	2	S48333	ORC1 protein - yea
29	35	68.6	2174	2	E95965	hypothetical glyci
30	35	68.6	2468	2	A83412	hypothetical prote
31	35	68.6	6642	2	T29757	protein UNC-89 - C
32	35	68.6	13055	2	T16580	hypothetical prote
33	34	66.7	88	2	S31030	gene 85 protein -
34	34	66.7	170	2	A97964	conserved hypothet
35	34	66.7	217	2	AI0987	probable lipoprote
36	34	66.7	228	2	G70532	hypothetical prote
37	34	66.7	273	2	T34234	hypothetical prote
38	34	66.7	275	2	D96926	prephenate dehydro
39	34	66.7	299	2	E82116	flagellar biosynth
40	34	66.7	353	2	T35221	probable ATP/GTP b
41	34	66.7	374	2	T46065	hypothetical prote
42	34	66.7	379	2	H70102	hypothetical prote
43	34	66.7	407	2	AF2497	transposase all715
44	34	66.7	424	2	T43498	hypothetical prote
45	34	66.7	443	2	D82975	two-component sens
46	34	66.7	544	2	T07593	pectinesterase (EC
47	34	66.7	550	2	S46528	pectinesterase (EC
48	34	66.7	630	2	T00352	hypothetical prote
49	34	66.7	678	2	T50256	probable vacuolar
50	34	66.7	686	2	A55665	microtubule-associ
51	34	66.7	813	2	G83662	class III stress r
52	34	66.7	1160	2	A46423	transcription fact
53	34	66.7	1176	2	T47444	hypothetical prote
54	34	66.7	1441	2	T39636	probable cleavage
55	34	66.7	2233	1	ZLNZP3	genome polyprotein
56	34	66.7	2340	2	B71704	cell surface antig
57	34	66.7	3643	2	T36410	probable polyketid
58	33	64.7	75	2	T12210	endopeptidase Clp
59	33	64.7	124	2	T24876	hypothetical prote
60	33	64.7	146	2	A69950	conserved hypothet
61	33	64.7	165	2	T26885	hypothetical prote
62	33	64.7	185	2	H82799	fimbrillin XF0487
63	33	64.7	194	2	E86885	hypothetical prote
64	33	64.7	219	2	AF0639	flagellar basal bo
65	33	64.7	237	2	JU0176	lectin alpha chain
66	33	64.7	280	2	S35103	bone sialoprotein
67	33	64.7	289	2	A89865	hypothetical prote
68	33	64.7	315	2	G91004	hypothetical prote
69	33	64.7	315	2	A85849	unknown protein en
70	33	64.7	341	2	E71564	probable cationic

```
 71       33      64.7       356    2   F84072              hypothetical prote
 72       33      64.7       366    2   E59102              hypothetical prote
 73       33      64.7       378    1   QXBY33              oxi3 intron 3 prot
 74       33      64.7       451    2   A86470              protein F21H2.12 [
 75       33      64.7       463    2   T14884              hypothetical prote
 76       33      64.7       466    2   E70865              trigger factor tig
 77       33      64.7       469    2   B87094              probable molecular
 78       33      64.7       513    2   S38197              sucrose transport
 79       33      64.7       585    2   C70330              conserved hypothet
 80       33      64.7       627    2   A41609              dnaK-type molecula
 81       33      64.7       656    2   D96831              hypothetical prote
 82       33      64.7       698    2   S52674              general sporulatio
 83       33      64.7       730    2   AI3480              penicillin-binding
 84       33      64.7       812    1   MMECOF              outer membrane ush
 85       33      64.7       843    2   S33442              EF protein - Strep
 86       33      64.7       923    1   B35905              endopeptidase Clp
 87       33      64.7       926    1   A35905              endopeptidase Clp
 88       33      64.7       967    2   S66852              hypothetical prote
 89       33      64.7       995    2   S50358              hypothetical prote
 90       33      64.7      1137    2   G70868              probable regulator
 91       33      64.7      1272    2   T30248              fragile X mental r
 92       33      64.7      1475    2   A44765              alpha-amylase (EC
 93       33      64.7      1822    2   S33441              EF protein - Strep
 94       33      64.7      1879    2   T19481              hypothetical prote
 95       33      64.7      1902    2   S06997              lactocepin (EC 3.4
 96       33      64.7      1902    2   B45764              lactocepin (EC 3.4
 97       33      64.7      3013    2   AB0480              probable invasin Y
 98       33      64.7      3283    2   AC1018              large repetitive p
 99       33      64.7      4447    2   A69679              polyketide synthas
100       33      64.7      4558    2   C82199              RTX toxin RtxA VC1
```

ALIGNMENTS

RESULT 1
CSBYC3
peptidylprolyl isomerase (EC 5.2.1.8) SCC3 precursor - yeast (Saccharomyces
cerevisiae)
N;Alternate names: cyclophilin SCC3; PPIase SCC3; protein YCR069w; protein
YCR070w
C;Species: Saccharomyces cerevisiae
C;Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C;Accession: S26658; S26587; S19484; S19517
R;Franco, L.; Jimenez, A.; Demolder, J.; Molemans, F.; Fiers, W.; Contreras, R.
Yeast 7, 971-979, 1991
A;Title: The nucleotide sequence of a third cyclophilin-homologous gene from
Saccharomyces cerevisiae.
A;Reference number: S26658; MUID:92206076; PMID:1803821
A;Accession: S26658
A;Molecule type: DNA
A;Residues: 1-318 <FRA>
A;Cross-references: UNIPROT:P25334; UNIPARC:UPI0000128C9D
R;Ballesta, J.P.G.; Franco, L.; Hoenicka, J.; Jimenez, A.; Remacha, M.; Sanz, E.
submitted to the Protein Sequence Database, October 1992
A;Reference number: S26587
A;Accession: S26587

A;Molecule type: DNA
A;Residues: 1-318 <BAL1>
A;Cross-references: UNIPARC:UPI0000128C9D; EMBL:X59720; NID:g1907116;
PIDN:CAA42275.1; PID:g1907209; GSPDB:GN00003; MIPS:YCR069w
A;Note: this is a revision to the sequence in reference S19482 and S19486
R;Contreras, R.; Demolder, J.; Fiers, W.; Molemans, F.
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19482
A;Accession: S19484
A;Molecule type: DNA
A;Residues: 1-170 <CON>
A;Cross-references: UNIPARC:UPI000017304A; EMBL:X59720; GSPDB:GN00003;
MIPS:YCR069w
A;Note: this sequence has been revised in reference S26587, resulting in
extension of the reading frame
R;Ballesta, J.P.G.; Franco, L.; Hoenicka, J.; ·Jimenez, A.; Remacha, M.; Sanz, E.
submitted to the Protein Sequence Database, March 1992
A;Reference number: S19486
A;Accession: S19517
A;Molecule type: DNA
A;Residues: 'MTGLKDSQWPILDLILTPRN',165-318 <BAL2>
A;Cross-references: UNIPARC:UPI000017304B; EMBL:X59720
A;Note: this was assumed to be protein YCR070w; the difference at the amino end
is due to a frameshift error
A;Note: this sequence has been revised in reference S26587
C;Genetics:
A;Gene: SGD:SCC3; MIPS:YCR069w
A;Cross-references: MIPS:YCR069w; SGD:S0000665
A;Map position: 3R
C;Function:
A;Description: catalyzes the cis-trans isomerization of peptidylproline peptide
bonds
C;Superfamily: peptidylprolyl isomerase SCC3; cyclophilin homology
C;Keywords: cis-trans-isomerase; glycoprotein; transmembrane protein
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-318/Product: peptidylprolyl isomerase SCC3 #status predicted <MAT>
F;51-279/Domain: cyclophilin homology <CYP>
F;286-303/Domain: transmembrane #status predicted <TMM>
F;166/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
   Query Match              78.4%;  Score 40;  DB 1;  Length 318;
   Best Local Similarity    70.0%;  Pred. No. 7.1;
   Matches     7;  Conservative    2;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 VITTDSDGNE 10
              :||| :||||
Db        171 IITTKADGNE 180
```

Search completed: December 8, 2005, 08:16:27
Job time : 24.3636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:16:37 ; Search time 5.90909 Seconds
                                              (without alignments)
                                              9.451 Million cell updates/sec

Title: US-10-789-494B-1
Perfect score: 51
Sequence: 1 VITTDSDGNE 10

Scoring table: BLOSUM62
                   Gapop 10.0 , Gapext 0.5

Searched: 32527 seqs, 5584426 residues

Total number of hits satisfying chosen parameters: 32527

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   Published_Applications_AA_New:*
                   1: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                   2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                   3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                   4: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                   5: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                   6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                   7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
                   8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                              SUMMARIES
                      %
Result   Query
   No.   Score   Match Length DB   ID   Description
----
     1   33   64.7   359 7   US-11-055-822-842   Sequence 842, App
     2   33   64.7   644 6   US-10-793-626-1436   Sequence 1436, Ap
     3   33   64.7   1170 6   US-10-858-730-71   Sequence 71, Appl
     4   32   62.7   378 6   US-10-467-657-5322   Sequence 5322, Ap
     5   32   62.7   740 7   US-11-110-837-2   Sequence 2, Appli
     6   32   62.7   740 7   US-11-110-837-4   Sequence 4, Appli
     7   32   62.7   784 6   US-10-467-657-5968   Sequence 5968, Ap
     8   31   60.8   399 6   US-10-467-657-7478   Sequence 7478, Ap
     9   31   60.8   1730 7   US-11-192-967-4   Sequence 4, Appli

10	31	60.8	1730	7	US-11-193-715-4	Sequence 4, Appli
11	31	60.8	2004	6	US-10-467-657-84	Sequence 84, Appl
12	31	60.8	2004	6	US-10-467-657-6322	Sequence 6322, Ap
13	31	60.8	2919	6	US-10-821-234-1133	Sequence 1133, Ap
14	30	58.8	261	6	US-10-467-657-6154	Sequence 6154, Ap
15	30	58.8	269	6	US-10-467-657-7278	Sequence 7278, Ap
16	30	58.8	318	6	US-10-821-234-1590	Sequence 1590, Ap
17	30	58.8	318	7	US-11-186-284-69	Sequence 69, Appl
18	30	58.8	438	7	US-11-140-417-2	Sequence 2, Appli
19	30	58.8	630	6	US-10-467-657-7014	Sequence 7014, Ap
20	30	58.8	710	7	US-11-089-551A-33	Sequence 33, Appl
21	29	56.9	141	6	US-10-793-626-2886	Sequence 2886, Ap
22	29	56.9	141	6	US-10-793-626-3084	Sequence 3084, Ap
23	29	56.9	168	6	US-10-793-626-2430	Sequence 2430, Ap
24	29	56.9	294	6	US-10-467-657-2876	Sequence 2876, Ap
25	29	56.9	319	6	US-10-793-626-792	Sequence 792, App
26	29	56.9	319	6	US-10-793-626-2008	Sequence 2008, Ap
27	29	56.9	325	7	US-11-074-176-368	Sequence 368, App
28	29	56.9	341	6	US-10-793-626-1278	Sequence 1278, Ap
29	29	56.9	373	6	US-10-131-826A-388	Sequence 388, App
30	29	56.9	421	6	US-10-858-730-1	Sequence 1, Appli
31	29	56.9	422	6	US-10-467-657-212	Sequence 212, App
32	29	56.9	422	6	US-10-467-657-6516	Sequence 6516, Ap
33	29	56.9	441	6	US-10-510-386-162	Sequence 162, App
34	29	56.9	463	6	US-10-793-626-960	Sequence 960, App
35	29	56.9	467	6	US-10-510-386-56	Sequence 56, Appl
36	29	56.9	486	6	US-10-793-626-788	Sequence 788, App
37	29	56.9	555	7	US-11-034-569-8	Sequence 8, Appli
38	29	56.9	560	7	US-11-034-569-6	Sequence 6, Appli
39	29	56.9	749	6	US-10-793-626-138	Sequence 138, App
40	29	56.9	817	6	US-10-793-626-50	Sequence 50, Appl
41	29	56.9	817	6	US-10-793-626-1528	Sequence 1528, Ap
42	29	56.9	926	6	US-10-841-129-2	Sequence 2, Appli
43	29	56.9	2314	7	US-11-013-759-11	Sequence 11, Appl
44	28.5	55.9	251	6	US-10-821-234-898	Sequence 898, App
45	28	54.9	96	7	US-11-084-554-217	Sequence 217, App
46	28	54.9	153	6	US-10-467-657-4546	Sequence 4546, Ap
47	28	54.9	193	6	US-10-467-657-2464	Sequence 2464, Ap
48	28	54.9	226	6	US-10-467-657-740	Sequence 740, App
49	28	54.9	312	7	US-11-055-822-34	Sequence 34, Appl
50	28	54.9	344	7	US-11-055-822-516	Sequence 516, App
51	28	54.9	403	6	US-10-528-031-7	Sequence 7, Appli
52	28	54.9	409	6	US-10-821-234-892	Sequence 892, App
53	28	54.9	419	7	US-11-084-624-18	Sequence 18, Appl
54	28	54.9	421	6	US-10-067-974-2	Sequence 2, Appli
55	28	54.9	421	6	US-10-067-974-16	Sequence 16, Appl
56	28	54.9	421	6	US-10-858-730-202	Sequence 202, App
57	28	54.9	421	6	US-10-525-710-73	Sequence 73, Appl
58	28	54.9	421	7	US-11-055-822-26	Sequence 26, Appl
59	28	54.9	452	7	US-11-074-176-274	Sequence 274, App
60	28	54.9	492	6	US-10-821-234-1108	Sequence 1108, Ap
61	28	54.9	496	6	US-10-485-517-324	Sequence 324, App
62	28	54.9	496	6	US-10-485-517-325	Sequence 325, App
63	28	54.9	578	6	US-10-821-234-1039	Sequence 1039, Ap
64	28	54.9	687	7	US-11-074-176-260	Sequence 260, App
65	28	54.9	741	6	US-10-467-657-6266	Sequence 6266, Ap
66	28	54.9	815	7	US-11-073-112-3	Sequence 3, Appli

```
67      28      54.9    890     7   US-11-106-623-28        Sequence 28, Appl
68      28      54.9    950     6   US-10-467-657-854       Sequence 854, App
69      28      54.9    1058    6   US-10-821-234-1473      Sequence 1473, Ap
70      28      54.9    1058    6   US-10-878-556A-63       Sequence 63, Appl
71      28      54.9    1058    7   US-11-069-642-105       Sequence 105, App
72      28      54.9    1377    6   US-10-467-657-7922      Sequence 7922, Ap
73      28      54.9    1394    6   US-10-467-657-7930      Sequence 7930, Ap
74      28      54.9    2376    7   US-11-096-051-4         Sequence 4, Appli
75      28      54.9    2657    6   US-10-821-234-1262      Sequence 1262, Ap
76      28      54.9    2715    7   US-11-096-051-2         Sequence 2, Appli
77      28      54.9    2715    7   US-11-113-424-51        Sequence 51, Appl
78      28      54.9    2721    7   US-11-096-051-10        Sequence 10, Appl
79      28      54.9    2725    7   US-11-096-051-8         Sequence 8, Appli
80      27      52.9    46      6   US-10-467-657-8850      Sequence 8850, Ap
81      27      52.9    123     7   US-11-076-713-3         Sequence 3, Appli
82      27      52.9    123     7   US-11-076-713-6         Sequence 6, Appli
83      27      52.9    154     6   US-10-467-657-158       Sequence 158, App
84      27      52.9    154     6   US-10-467-657-6520      Sequence 6520, Ap
85      27      52.9    154     6   US-10-467-657-7634      Sequence 7634, Ap
86      27      52.9    171     7   US-11-055-822-352       Sequence 352, App
87      27      52.9    179     6   US-10-793-626-1738      Sequence 1738, Ap
88      27      52.9    187     6   US-10-467-657-1636      Sequence 1636, Ap
89      27      52.9    198     6   US-10-467-657-4664      Sequence 4664, Ap
90      27      52.9    201     6   US-10-467-657-2006      Sequence 2006, Ap
91      27      52.9    227     6   US-10-467-657-2624      Sequence 2624, Ap
92      27      52.9    304     6   US-10-467-657-7410      Sequence 7410, Ap
93      27      52.9    320     7   US-11-000-463-354       Sequence 354, App
94      27      52.9    344     6   US-10-878-556A-123      Sequence 123, App
95      27      52.9    351     6   US-10-467-657-8316      Sequence 8316, Ap
96      27      52.9    373     6   US-10-467-657-5582      Sequence 5582, Ap
97      27      52.9    409     6   US-10-467-657-4220      Sequence 4220, Ap
98      27      52.9    409     6   US-10-467-657-7088      Sequence 7088, Ap
99      27      52.9    418     6   US-10-793-626-1288      Sequence 1288, Ap
100     27      52.9    434     7   US-11-055-822-794       Sequence 794, App
```

ALIGNMENTS

RESULT 1
US-11-055-822-842
; Sequence 842, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
.; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS
; FILE REFERENCE: BGI-121CPCN
; CURRENT APPLICATION NUMBER: US/11/055,822
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: 09/606,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031

; SEQ ID NO 842
;   LENGTH: 359
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-11-055-822-842

  Query Match   64.7%;   Score 33;   DB 7;   Length 359;
  Best Local Similarity   60.0%;   Pred. No. 23;
  Matches   6; Conservative   2; Mismatches   2; Indels   0; Gaps   0;

Qy   1 VITTDSDGNE 10
             | ||:|||:
Db   78 VSLTDADGND 87

Search completed: December 8, 2005, 08:35:41
Job time : 6.90909 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:15:47 ; Search time 101.364 Seconds
                                            (without alignments)
                                            41.221 Million cell updates/sec

Title: US-10-789-494B-1
Perfect score: 51
Sequence: 1 VITTDSDGNE 10

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   Published_Applications_AA_Main:*
                 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
                 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                    SUMMARIES
                          %
Result   Query
   No.   Score   Match Length DB   ID   Description
----
        1   51 100.0   10 5   US-10-789-494B-1   Sequence 1, Appli
        2   51 100.0   151 5   US-10-789-494B-9   Sequence 9, Appli
        3   43   84.3   317 6   US-11-097-143-30096   Sequence 30096, A
        4   39   76.5   475 6   US-11-097-143-13803   Sequence 13803, A
        5   39   76.5   534 4   US-10-046-232-22   Sequence 22, Appl
        6   39   76.5   534 5   US-10-940-954-22   Sequence 22, Appl
        7   39   76.5   559 4   US-10-046-232-20   Sequence 20, Appl
        8   39   76.5   559 5   US-10-940-954-20   Sequence 20, Appl
        9   39   76.5   636 4   US-10-408-765A-2973   Sequence 2973, Ap
       10   38   74.5   437 4   US-10-369-493-6306   Sequence 6306, Ap
       11   38   74.5   437 4   US-10-602-268-21   Sequence 21, Appl

12	38	74.5	450	4	US-10-148-884-4	Sequence 4, Appli
13	38	74.5	608	4	US-10-148-884-2	Sequence 2, Appli
14	38	74.5	953	3	US-09-884-696-3	Sequence 3, Appli
15	38	74.5	953	4	US-10-148-884-5	Sequence 5, Appli
16	37	72.5	129	3	US-09-731-660A-3	Sequence 3, Appli
17	37	72.5	508	3	US-09-731-660A-1	Sequence 1, Appli
18	37	72.5	508	3	US-09-757-415A-1	Sequence 1, Appli
19	37	72.5	508	4	US-10-146-473-67	Sequence 67, Appl
20	37	72.5	521	4	US-10-276-774-2192	Sequence 2192, Ap
21	37	72.5	862	3	US-09-738-626-3956	Sequence 3956, Ap
22	37	72.5	862	5	US-10-494-672-308	Sequence 308, App
23	37	72.5	942	4	US-10-437-963-198719	Sequence 198719,
24	37	72.5	1184	4	US-10-437-963-198716	Sequence 198716,
25	36	70.6	113	4	US-10-437-963-181695	Sequence 181695,
26	36	70.6	212	5	US-10-501-282-5738	Sequence 5738, Ap
27	36	70.6	230	3	US-09-793-708-17	Sequence 17, Appl
28	36	70.6	230	4	US-10-201-365-21	Sequence 21, Appl
29	36	70.6	230	4	US-10-160-539-17	Sequence 17, Appl
30	36	70.6	230	5	US-10-468-828-17	Sequence 17, Appl
31	36	70.6	230	5	US-10-846-335-17	Sequence 17, Appl
32	36	70.6	256	5	US-10-501-282-5740	Sequence 5740, Ap
33	36	70.6	781	4	US-10-282-122A-49736	Sequence 49736, A
34	36	70.6	924	5	US-10-450-763-58411	Sequence 58411, A
35	36	70.6	1481	4	US-10-050-763-1	Sequence 1, Appli
36	36	70.6	2402	4	US-10-661-809-20	Sequence 20, Appl
37	35	68.6	70	4	US-10-425-115-225379	Sequence 225379,
38	35	68.6	86	4	US-10-425-115-225376	Sequence 225376,
39	35	68.6	96	4	US-10-425-115-246405	Sequence 246405,
40	35	68.6	112	4	US-10-425-115-225380	Sequence 225380,
41	35	68.6	177	5	US-10-660-811A-184	Sequence 184, App
42	35	68.6	202	4	US-10-437-963-106705	Sequence 106705,
43	35	68.6	324	4	US-10-320-797-3094	Sequence 3094, Ap
44	35	68.6	364	4	US-10-282-122A-59693	Sequence 59693, A
45	35	68.6	475	4	US-10-369-493-12784	Sequence 12784, A
46	35	68.6	498	4	US-10-424-599-196154	Sequence 196154,
47	35	68.6	587	4	US-10-425-115-283525	Sequence 283525,
48	35	68.6	625	4	US-10-661-809-19	Sequence 19, Appl
49	35	68.6	772	6	US-11-097-143-31401	Sequence 31401, A
50	35	68.6	833	4	US-10-282-122A-52796	Sequence 52796, A
51	35	68.6	899	6	US-11-097-143-23256	Sequence 23256, A
52	35	68.6	899	6	US-11-097-143-23259	Sequence 23259, A
53	35	68.6	914	4	US-10-369-493-1851	Sequence 1851, Ap
54	35	68.6	1132	5	US-10-732-923-3315	Sequence 3315, Ap
55	35	68.6	1166	5	US-10-732-923-3316	Sequence 3316, Ap
56	35	68.6	1572	4	US-10-282-122A-69415	Sequence 69415, A
57	35	68.6	2468	4	US-10-246-330-4	Sequence 4, Appli
58	35	68.6	2468	4	US-10-282-122A-66335	Sequence 66335, A
59	35	68.6	4025	4	US-10-437-963-193926	Sequence 193926,
60	35	68.6	6642	4	US-10-369-493-5013	Sequence 5013, Ap
61	34	66.7	41	4	US-10-437-963-106353	Sequence 106353,
62	34	66.7	42	4	US-10-424-599-221238	Sequence 221238,
63	34	66.7	103	5	US-10-473-757-2	Sequence 2, Appli
64	34	66.7	119	4	US-10-437-963-204546	Sequence 204546,
65	34	66.7	155	4	US-10-437-963-171304	Sequence 171304,
66	34	66.7	199	4	US-10-437-963-175368	Sequence 175368,
67	34	66.7	201	4	US-10-425-114-56602	Sequence 56602, A
68	34	66.7	217	4	US-10-282-122A-72946	Sequence 72946, A

```
 69      34      66.7     217    4   US-10-282-122A-75502      Sequence 75502, A
 70      34      66.7     228    3   US-09-791-171-66          Sequence 66, Appl
 71      34      66.7     228    3   US-09-804-980-66          Sequence 66, Appl
 72      34      66.7     228    4   US-10-620-246-66          Sequence 66, Appl
 73      34      66.7     240    4   US-10-425-115-271893      Sequence 271893,
 74      34      66.7     263    6   US-11-097-143-36729       Sequence 36729, A
 75      34      66.7     264    4   US-10-263-367-6           Sequence 6, Appli
 76      34      66.7     352    5   US-10-994-726-78          Sequence 78, Appl
 77      34      66.7     354    5   US-10-450-763-48490       Sequence 48490, A
 78      34      66.7     364    5   US-10-450-763-44750       Sequence 44750, A
 79      34      66.7     378    4   US-10-437-963-120806      Sequence 120806,
 80      34      66.7     379    5   US-10-994-726-77          Sequence 77, Appl
 81      34      66.7     403    4   US-10-282-122A-61338      Sequence 61338, A
 82      34      66.7     438    4   US-10-437-963-168400      Sequence 168400,
 83      34      66.7     499    5   US-10-739-930-8216        Sequence 8216, Ap
 84      34      66.7     501    4   US-10-029-386-32319       Sequence 32319, A
 85      34      66.7     570    5   US-10-732-923-8437        Sequence 8437, Ap
 86      34      66.7     707    4   US-10-735-098-10          Sequence 10, Appl
 87      34      66.7     813    4   US-10-369-493-17101       Sequence 17101, A
 88      34      66.7     813    5   US-10-732-923-7201        Sequence 7201, Ap
 89      34      66.7     953    6   US-11-097-143-18594       Sequence 18594, A
 90      34      66.7     973    4   US-10-276-774-2310        Sequence 2310, Ap
 91      34      66.7    1285    5   US-10-450-763-46696       Sequence 46696, A
 92      34      66.7    1391    4   US-10-437-963-128235      Sequence 128235,
 93      34      66.7    1632    4   US-10-282-122A-49890      Sequence 49890, A
 94      34      66.7    1849    4   US-10-637-544-2           Sequence 2, Appli
 95      34      66.7    1849    5   US-10-819-275-2           Sequence 2, Appli
 96      34      66.7    2233    5   US-10-789-400-14          Sequence 14, Appl
 97      33      64.7      59    4   US-10-425-115-219643      Sequence 219643,
 98      33      64.7      66    3   US-09-867-550-728         Sequence 728, App
 99      33      64.7      89    4   US-10-424-599-205325      Sequence 205325,
100      33      64.7      93    4   US-10-424-599-245886      Sequence 245886,
```

ALIGNMENTS

RESULT 1
US-10-789-494B-1
; Sequence 1, Application US/10789494B
; Publication No. US20050143296A1
; GENERAL INFORMATION:
;   APPLICANT: TSUBOUCHI, Kozo
;   APPLICANT: YAMADA, Hiromi
;   TITLE OF INVENTION: EXTRACTION AND UTILIZATION OF CELL
;   TITLE OF INVENTION: GROWTH-PROMOTING PEPTIDES FROM SILK PROTEIN
;   FILE REFERENCE: OPS 635
;   CURRENT APPLICATION NUMBER: US/10/789,494B
;   CURRENT FILING DATE: 2004-02-27
;   PRIOR APPLICATION NUMBER: JP 2003-55048
;   PRIOR FILING DATE: 2003-02-28
;   NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 1
;   LENGTH: 10
;   TYPE: PRT
;   ORGANISM: Bombyx mori
US-10-789-494B-1

```
Query Match           100.0%;  Score 51;  DB 5;  Length 10;
Best Local Similarity 100.0%;  Pred. No. 0.017;
Matches  10;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;


Qy        1 VITTDSDGNE 10
            ||||||||||
Db        1 VITTDSDGNE 10


RESULT 3
US-11-097-143-30096
; Sequence 30096, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
;  APPLICANT: Venter, J. Craig
;  APPLICANT:  et al.
;  TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
;  TITLE OF INVENTION:  ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
;  TITLE OF INVENTION:  DROSOPHILA GENES.
;  FILE REFERENCE: CL000728
;  CURRENT APPLICATION NUMBER: US/11/097,143
;  CURRENT FILING DATE:  2005-04-04
;  PRIOR APPLICATION NUMBER: 60/157,832
;  PRIOR FILING DATE: 1999-10-05
;  PRIOR APPLICATION NUMBER: 60/160,191
;  PRIOR FILING DATE: 1999-10-19
;  PRIOR APPLICATION NUMBER: 60/161,932
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: 60/164,769
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/173,383
;  PRIOR FILING DATE: 1999-12-28
;  PRIOR APPLICATION NUMBER: 60/175,693
;  PRIOR FILING DATE: 2000-01-12
;  PRIOR APPLICATION NUMBER: 60/184,831
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: 60/191,637
;  PRIOR FILING DATE: 2000-03-23
;  NUMBER OF SEQ ID NOS: 43008
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 30096
;   LENGTH: 317
;   TYPE: PRT
;   ORGANISM: DROSOPHILA
US-11-097-143-30096

Query Match            84.3%;  Score 43;  DB 6;  Length 317;
Best Local Similarity  88.9%;  Pred. No. 21;
Matches  8;  Conservative  1;  Mismatches  0;  Indels  0;  Gaps  0;


Qy        2 ITTDSDGNE 10
            ||:|||||
Db      141 ITSDSDGNE 149


Search completed: December  8, 2005, 08:35:24
Job time : 104.364 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:07:46 ; Search time 30 Seconds
                                              (without alignments)
                                              27.559 Million cell updates/sec

Title: US-10-789-494B-1
Perfect score: 51
Sequence: 1 VITTDSDGNE 10

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   Issued_Patents_AA:*
                 1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
                 2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
                 3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
                 4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
                 5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
                 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

       Pred. No. is the number of results predicted by chance to have a
       score greater than or equal to the score of the result being printed,
       and is derived by analysis of the total score distribution.

                                    SUMMARIES

		% Query				
Result No.	Score	Match	Length	DB	ID	Description
1	39	76.5	534	2	US-10-046-232-22	Sequence 22, Appl
2	39	76.5	559	2	US-10-046-232-20	Sequence 20, Appl
3	38	74.5	924	2	US-08-619-812-8	Sequence 8, Appli
4	38	74.5	926	1	US-07-908-253-2	Sequence 2, Appli
5	38	74.5	926	1	US-08-455-970A-2	Sequence 2, Appli
6	38	74.5	926	1	US-08-387-156-6	Sequence 6, Appli
7	38	74.5	926	1	US-08-694-865-6	Sequence 6, Appli
8	38	74.5	926	1	US-08-878-748-6	Sequence 6, Appli
9	38	74.5	926	1	US-08-535-837-2	Sequence 2, Appli
10	38	74.5	926	2	US-09-124-491-6	Sequence 6, Appli
11	38	74.5	926	2	US-09-383-912-6	Sequence 6, Appli

12	38	74.5	926	2	US-08-976-566-2	Sequence 2, Appli
13	38	74.5	926	6	5476657-3	Patent No. 5476657
14	38	74.5	936	1	US-08-455-970A-12	Sequence 12, Appl
15	38	74.5	936	2	US-08-976-566-12	Sequence 12, Appl
16	38	74.5	943	1	US-08-455-970A-10	Sequence 10, Appl
17	38	74.5	943	2	US-08-976-566-10	Sequence 10, Appl
18	38	74.5	951	1	US-08-455-970A-14	Sequence 14, Appl
19	38	74.5	951	2	US-08-976-566-14	Sequence 14, Appl
20	38	74.5	977	1	US-08-387-156-8	Sequence 8, Appli
21	38	74.5	977	1	US-08-694-865-8	Sequence 8, Appli
22	38	74.5	977	1	US-08-878-748-8	Sequence 8, Appli
23	38	74.5	977	2	US-09-124-491-8	Sequence 8, Appli
24	38	74.5	977	2	US-09-383-912-8	Sequence 8, Appli
25	38	74.5	1069	1	US-07-777-715-9	Sequence 9, Appli
26	38	74.5	1069	1	US-08-170-126-4	Sequence 4, Appli
27	38	74.5	1069	2	US-08-954-418-4	Sequence 4, Appli
28	38	74.5	1098	1	US-07-777-715-7	Sequence 7, Appli
29	38	74.5	1098	1	US-08-170-126-2	Sequence 2, Appli
30	38	74.5	1098	2	US-08-954-418-2	Sequence 2, Appli
31	37	72.5	129	2	US-08-980-523-11	Sequence 11, Appl
32	37	72.5	508	2	US-08-980-523-9	Sequence 9, Appli
33	37	72.5	729	2	US-09-543-681A-8257	Sequence 8257, Ap
34	37	72.5	856	2	US-09-605-703B-2760	Sequence 2760, Ap
35	37	72.5	3290	2	US-09-328-352-5486	Sequence 5486, Ap
36	36	70.6	230	2	US-09-320-878-17	Sequence 17, Appl
37	36	70.6	230	2	US-09-141-908-21	Sequence 21, Appl
38	36	70.6	230	2	US-09-657-440-17	Sequence 17, Appl
39	36	70.6	230	2	US-09-793-708-17	Sequence 17, Appl
40	36	70.6	1481	2	US-10-050-763-1	Sequence 1, Appli
41	35	68.6	82	2	US-09-248-796A-20781	Sequence 20781, A
42	35	68.6	161	2	US-09-538-092-756	Sequence 756, App
43	35	68.6	209	2	US-09-252-991A-26544	Sequence 26544, A
44	35	68.6	332	2	US-09-248-796A-18352	Sequence 18352, A
45	35	68.6	403	2	US-09-489-039A-11910	Sequence 11910, A
46	35	68.6	664	2	US-09-107-532A-7252	Sequence 7252, Ap
47	35	68.6	914	1	US-08-484-105-2	Sequence 2, Appli
48	35	68.6	914	1	US-08-484-106-2	Sequence 2, Appli
49	35	68.6	2736	2	US-09-252-991A-30227	Sequence 30227, A
50	34	66.7	64	2	US-09-248-796A-22857	Sequence 22857, A
51	34	66.7	228	2	US-09-050-739-66	Sequence 66, Appl
52	34	66.7	301	2	US-09-252-991A-24016	Sequence 24016, A
53	34	66.7	339	2	US-09-543-681A-7617	Sequence 7617, Ap
54	34	66.7	352	2	US-09-830-230A-78	Sequence 78, Appl
55	34	66.7	379	2	US-09-830-230A-77	Sequence 77, Appl
56	33	64.7	39	2	US-09-286-691-19	Sequence 19, Appl
57	33	64.7	39	2	US-09-687-147-19	Sequence 19, Appl
58	33	64.7	182	2	US-09-540-236-2367	Sequence 2367, Ap
59	33	64.7	213	2	US-09-902-540-14656	Sequence 14656, A
60	33	64.7	308	2	US-09-248-796A-24139	Sequence 24139, A
61	33	64.7	345	2	US-09-107-532A-7021	Sequence 7021, Ap
62	33	64.7	366	2	US-10-172-527A-18	Sequence 18, Appl
63	33	64.7	377	2	US-09-248-796A-20227	Sequence 20227, A
64	33	64.7	381	2	US-10-052-092-29	Sequence 29, Appl
65	33	64.7	385	2	US-09-248-796A-21836	Sequence 21836, A
66	33	64.7	463	1	US-08-853-659A-52	Sequence 52, Appl
67	33	64.7	491	2	US-09-807-258-18	Sequence 18, Appl
68	33	64.7	644	2	US-09-710-279-1436	Sequence 1436, Ap

```
69      33    64.7     669    2   US-09-071-035-264        Sequence 264, App
70      33    64.7     669    2   US-10-206-576-264        Sequence 264, App
71      33    64.7     698    2   US-09-538-092-151        Sequence 151, App
72      33    64.7     716    1   US-08-372-652-4          Sequence 4, Appli
73      33    64.7     716    4   PCT-US95-16311-4         Sequence 4, Appli
74      33    64.7     849    2   US-09-949-016-9522       Sequence 9522, Ap
75      33    64.7    1638    2   US-09-071-035-258        Sequence 258, App
76      33    64.7    1638    2   US-09-071-035-262        Sequence 262, App
77      33    64.7    1638    2   US-09-071-035-266        Sequence 266, App
78      33    64.7    1638    2   US-10-206-576-258        Sequence 258, App
79      33    64.7    1638    2   US-10-206-576-262        Sequence 262, App
80      33    64.7    1638    2   US-10-206-576-266        Sequence 266, App
81      33    64.7    1747    2   US-09-134-000C-5999      Sequence 5999, Ap
82      33    64.7    2233    1   US-08-569-853-1          Sequence 1, Appli
83      33    64.7    2233    1   US-08-569-853-2          Sequence 2, Appli
84      33    64.7    2233    2   US-08-987-439-1          Sequence 1, Appli
85      33    64.7   10182    2   US-09-134-001C-3159      Sequence 3159, Ap
86      32    62.7     136    2   US-09-270-767-32469      Sequence 32469, A
87      32    62.7     136    2   US-09-270-767-47686      Sequence 47686, A
88      32    62.7     148    2   US-09-248-528-18         Sequence 18, Appl
89      32    62.7     148    2   US-09-549-108-18         Sequence 18, Appl
90      32    62.7     148    2   US-09-549-111-18         Sequence 18, Appl
91      32    62.7     148    2   US-09-549-106-18         Sequence 18, Appl
92      32    62.7     148    2   US-09-550-394-18         Sequence 18, Appl
93      32    62.7     160    2   US-09-605-703B-1532      Sequence 1532, Ap
94      32    62.7     181    2   US-09-543-681A-4179      Sequence 4179, Ap
95      32    62.7     217    2   US-09-252-991A-31584     Sequence 31584, A
96      32    62.7     234    2   US-09-902-540-13573      Sequence 13573, A
97      32    62.7     246    2   US-09-270-767-44993      Sequence 44993, A
98      32    62.7     254    2   US-09-489-039A-10109     Sequence 10109, A
99      32    62.7     281    2   US-09-107-532A-4658      Sequence 4658, Ap
100     32    62.7     366    2   US-09-605-703B-1530      Sequence 1530, Ap
```

ALIGNMENTS

RESULT 1
US-10-046-232-22
; Sequence 22, Application US/10046232
; Patent No. 6861243
; GENERAL INFORMATION:
; APPLICANT: Helmut SCHWAB
; APPLICANT: Anton GLIEDER
; APPLICANT: Christoph KRATKY
; APPLICANT: Ingrid DREVENY
; APPLICANT: Peter POCHLAUER
; APPLICANT: Wolfgang SKRANC
; APPLICANT: Herbert MAYRHOFER
; APPLICANT: Irma WIRTH
; APPLICANT: Rudolf NEUHOFER
; APPLICANT: Rodolfo BONA
; TITLE OF INVENTION: New genes containing a DNA sequence coding for a
hydroxynitrile lyase,
; TITLE OF INVENTION: recombinant proteins derived therefrom and having
hydroxynitrile lyase activity, and use
; TITLE OF INVENTION: thereof

    Query Match   76.5%; Score 39; DB 2; Length 534;
    Best Local Similarity   88.9%; Pred. No. 63;
    Matches   8; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 VITTDSDGN 9
              || ||||||
Db   247 VIYTDSDGN 255

Search completed: December 8, 2005, 08:17:37
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:03:20 ; Search time 68.6364 Seconds
                                                (without alignments)
                                              64.015 Million cell updates/sec

Title: US-10-789-494B-1
Perfect score: 51
Sequence: 1 VITTDSDGNE 10

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   A_Geneseq_21:*
                 1: geneseqp1980s:*
                 2: geneseqp1990s:*
                 3: geneseqp2000s:*
                 4: geneseqp2001s:*
                 5: geneseqp2002s:*
                 6: geneseqp2003as:*
                 7: geneseqp2003bs:*
                 8: geneseqp2004s:*
                 9: geneseqp2005s:*

          Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.

                                 SUMMARIES

         %
Result   Query
  No.   Score Match Length DB ID   Description
----
      1   51 100.0   10 8 ADU51205   Adu51205 Silkworm
      2   51 100.0   151 8 ADU51163   Adu51163 Domestic
      3   43   84.3   317 4 ABB67768   Abb67768 Drosophil
      4   40   78.4   101 5 ABP03713   Abp03713 Human ORF
      5   39   76.5   475 4 ABB62337   Abb62337 Drosophil
      6   39   76.5   532 8 ADO60398   Ado60398 (R)-hydro
      7   39   76.5   534 5 AAB71495   Aab71495 P. amygda
      8   39   76.5   559 5 AAB71494   Aab71494 Almond md

9	39	76.5	636	7	ADJ71167	Adj71167	Human hea
10	38	74.5	252	5	ABP29176	Abp29176	Streptoco
11	38	74.5	437	8	ADN23653	Adn23653	Bacterial
12	38	74.5	450	4	AAE04637	Aae04637	Pasteurel
13	38	74.5	608	4	AAE04636	Aae04636	Pasteurel
14	38	74.5	617	4	AAU39683	Aau39683	Propionib
15	38	74.5	617	6	ABM36202	Abm36202	Propionib
16	38	74.5	924	2	AAR10889	Aar10889	Leukotoxi
17	38	74.5	924	2	AAR42378	Aar42378	Recombina
18	38	74.5	924	2	AAR42380	Aar42380	Recombina
19	38	74.5	924	2	AAR42385	Aar42385	Recombina
20	38	74.5	926	2	AAR14482	Aar14482	LKT352. 1
21	38	74.5	926	2	AAR34545	Aar34545	Leukotoxi
22	38	74.5	926	2	AAR50291	Aar50291	Recombina
23	38	74.5	926	2	AAW03945	Aaw03945	P. haemol
24	38	74.5	926	2	AAW79568	Aaw79568	Leukotoxi
25	38	74.5	936	2	AAR34547	Aar34547	GnRH-leuk
26	38	74.5	943	2	AAR34546	Aar34546	Somatosta
27	38	74.5	951	2	AAR34548	Aar34548	Rotavirus
28	38	74.5	953	2	AAR07167	Aar07167	105kD PTX
29	38	74.5	953	2	AAR15159	Aar15159	Leukotoxi
30	38	74.5	953	2	AAR43865	Aar43865	Leukotoxi
31	38	74.5	953	2	AAR60072	Aar60072	PtxA prot
32	38	74.5	953	4	AAE04638	Aae04638	Pasteurel
33	38	74.5	977	2	AAW03942	Aaw03942	LKT-GnRH
34	38	74.5	977	2	AAW79569	Aaw79569	LKT-GnRH
35	38	74.5	1069	2	AAR52748	Aar52748	Bovine IF
36	38	74.5	1069	2	AAW13867	Aaw13867	Chimeric
37	38	74.5	1069	3	AAB21074	Aab21074	Bovine ga
38	38	74.5	1098	2	AAR22103	Aar22103	Bovine IL
39	38	74.5	1098	2	AAR52747	Aar52747	Bovine IL
40	38	74.5	1098	2	AAW13866	Aaw13866	Chimeric
41	38	74.5	1098	3	AAB21073	Aab21073	Bovine IL
42	37	72.5	101	2	AAY07018	Aay07018	Breast ca
43	37	72.5	508	2	AAW62558	Aaw62558	Fibroblas
44	37	72.5	508	4	AAU04693	Aau04693	Human suc
45	37	72.5	508	7	ADC35101	Adc35101	Human bre
46	37	72.5	521	4	ABB11822	Abb11822	Human FGF
47	37	72.5	729	7	ADF07972	Adf07972	Bacterial
48	37	72.5	862	4	AAG90202	Aag90202	C glutami
49	37	72.5	862	7	ADL65951	Adl65951	C. glutam
50	37	72.5	1482	6	ABR58656	Abr58656	Human can
51	37	72.5	1497	5	ABP69627	Abp69627	Human pol
52	37	72.5	3290	6	ADA34199	Ada34199	Acinetoba
53	36	70.6	212	6	ADB11322	Adb11322	Alloiococ
54	36	70.6	230	3	AAB18653	Aab18653	Amino aci
55	36	70.6	230	3	AAY67217	Aay67217	ORF 16 en
56	36	70.6	230	6	ABG71677	Abg71677	Partial p
57	36	70.6	230	6	ADA09416	Ada09416	S. venezu
58	36	70.6	230	7	ADH53460	Adh53460	Streptomy
59	36	70.6	256	6	ADB11320	Adb11320	Alloiococ
60	36	70.6	781	6	ABU21812	Abu21812	Protein e
61	36	70.6	924	4	ABG28052	Abg28052	Novel hum
62	36	70.6	1122	2	AAR64927	Aar64927	Cytadhesi
63	36	70.6	1481	8	ADM47534	Adm47534	Thermoana
64	36	70.6	2402	8	ADO84848	Ado84848	S epiderm
65	35	68.6	161	7	ADK64272	Adk64272	Disease t

```
  66      35    68.6     177   8   ADL81912              Adl81912 P. aerugi
  67      35    68.6     209   7   ABO77798              Abo77798 Pseudomon
  68      35    68.6     324   7   ADB70050              Adb70050 C. neofor
  69      35    68.6     364   6   ABU31769              Abu31769 Protein e
  70      35    68.6     403   7   ABO65393              Abo65393 Klebsiell
  71      35    68.6     475   8   ADS23751              Ads23751 Bacterial
  72      35    68.6     625   8   ADO84895              Ado84895 E faecium
  73      35    68.6     625   9   ADV16668              Adv16668 E. faeciu
  74      35    68.6     664   7   ADC97625              Adc97625 E. faeciu
  75      35    68.6     772   4   ABB68203              Abb68203 Drosophil
  76      35    68.6     833   6   ABU24872              Abu24872 Protein e
  77      35    68.6     899   4   ABB65489              Abb65489 Drosophil
  78      35    68.6     899   4   ABB65488              Abb65488 Drosophil
  79      35    68.6     914   2   AAR77274              Aar77274 ORC1 subu
  80      35    68.6     914   2   AAW22224              Aaw22224 S. cerevi
  81      35    68.6     914   6   ABR53642              Abr53642 Protein s
  82      35    68.6     914   7   ADK64132              Adk64132 Disease t
  83      35    68.6     914   8   ADN19198              Adn19198 Bacterial
  84      35    68.6    1572   6   ABU41491              Abu41491 Protein e
  85      35    68.6    2468   6   ABU38411              Abu38411 Protein e
  86      35    68.6    2468   6   ABP59933              Abp59933 Microbial
  87      35    68.6    2736   7   ABO81481              Abo81481 Pseudomon
  88      35    68.6    6642   8   ADN22360              Adn22360 Bacterial
  89      34    66.7      55   3   AAB34703              Aab34703 Human sec
  90      34    66.7      79   3   AAG33475              Aag33475 Arabidops
  91      34    66.7     103   5   ABG72805              Abg72805 Human cyt
  92      34    66.7     177   7   ADM25480              Adm25480 Hyperther
  93      34    66.7     188   7   ABM89977              Abm89977 Rice abio
  94      34    66.7     201   8   ADX93938              Adx93938 Plant ful
  95      34    66.7     217   6   ABU47578              Abu47578 Protein e
  96      34    66.7     217   6   ABU45022              Abu45022 Protein e
  97      34    66.7     228   2   AAW72909              Aaw72909 Mycobacte
  98      34    66.7     228   2   AAY21926              Aay21926 Amino aci
  99      34    66.7     263   4   ABB69979              Abb69979 Drosophil
 100      34    66.7     301   7   ABO75270              Abo75270 Pseudomon
```

ALIGNMENTS

RESULT 1
ADU51205
ID   ADU51205 standard; peptide; 10 AA.
XX
AC   ADU51205;
XX
DT   24-FEB-2005   (first entry)
XX
DE   Silkworm fibroin-derived fibroblast proliferation peptide 2.
XX
KW   vulnerary; cell proliferation; wound healing; cell adhesion; cosmetics;
KW   cell culture; fibroin.
XX
OS   Bombycoidea.
OS   Synthetic.
XX
PN   JP2004339189-A.

```
   Query Match             100.0%;  Score 51;  DB 8;  Length 10;
   Best Local Similarity   100.0%;  Pred. No. 0.024;
   Matches   10;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         1 VITTDSDGNE 10
             ||||||||||
Db         1 VITTDSDGNE 10
```

Search completed: December 8, 2005, 08:10:37
Job time : 73.6364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:04:00 ; Search time 159.273 Seconds
                                          (without alignments)
                                          53.156 Million cell updates/sec

Title: US-10-789-494B-5
Perfect score: 75
Sequence: 1 YGWGDGGYGSDS 12

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   UniProt_05.80:*
                 1: uniprot_sprot:*
                 2: uniprot_trembl:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                  SUMMARIES
                      %
Result   Query
   No.   Score Match Length DB   ID   Description
   ----
      1   75   100.0   421   2   Q93119_ANTPE   Q93119 antheraea p
      2   75   100.0   436   2   Q967T8_ANTPE   Q967t8 antheraea p
      3   75   100.0   507   2   Q8ISB3_9NEOP   Q8isb3 antheraea m
      4   75   100.0   2639   2   O76786_ANTPE   O76786 antheraea p
      5   75   100.0   2655   2   Q964F4_ANTYA   Q964f4 antheraea y
      6   55   73.3   151   2   Q95VQ0_ANTYA   Q95vq0 antheraea y
      7   53   70.7   445   2   Q4X6C2_PLACH   Q4x6c2 plasmodium
      8   53   70.7   479   2   Q7RKG8_PLAYO   Q7rkg8 plasmodium
      9   53   70.7   479   2   Q4Z184_PLABE   Q4z184 plasmodium
     10   53   70.7   480   2   Q8IDD6_PLAF7   Q8idd6 plasmodium
     11   52.5   70.0   142   2   Q97LH4_CLOAB   Q97lh4 clostridium
     12   52   69.3   1410   2   Q8CMJ0_SHEON   Q8cmj0 shewanella
     13   52   69.3   1422   2   Q8EFU3_SHEON   Q8efu3 shewanella
     14   51   68.0   169   1   GRP10_BRANA   Q05966 brassica na
     15   50   66.7   599   2   Q82M54_STRAW   Q82m54 streptomyce

16	50	66.7	602	2	O87849_STRCO	O87849 streptomyce
17	50	66.7	682	2	Q87GL4_VIBPA	Q87gl4 vibrio para
18	50	66.7	848	2	Q9RK65_STRCO	Q9rk65 streptomyce
19	50	66.7	936	2	Q4I9I0_GIBZE	Q4i9i0 gibberella
20	50	66.7	1009	2	Q6CHC1_YARLI	Q6chc1 yarrowia li
21	50	66.7	1326	2	Q4Q8G0_LEIMA	Q4q8g0 leishmania
22	48	64.0	76	2	Q6FDZ6_ACIAD	Q6fdz6 acinetobact
23	48	64.0	172	2	Q953P2_9PSIT	Q953p2 amazona och
24	48	64.0	214	1	GRP2_NICSY	P27484 nicotiana s
25	48	64.0	331	2	Q8KXI0_SHIFL	Q8kxi0 shigella fl
26	47	62.7	92	2	O24350_SILLA	O24350 silene lati
27	47	62.7	374	2	Q6LPG6_PHOPR	Q6lpg6 photobacter
28	47	62.7	443	2	Q7EZ34_ORYSA	Q7ez34 oryza sativ
29	47	62.7	527	2	Q5KFM5_CRYNE	Q5kfm5 cryptococcu
30	47	62.7	546	2	Q55QJ1_CRYNE	Q55qj1 cryptococcu
31	47	62.7	1013	2	Q97C94_THEVO	Q97c94 thermoplasm
32	47	62.7	1022	2	Q82MA7_STRAW	Q82ma7 streptomyce
33	47	62.7	1408	2	Q8E833_SHEON	Q8e833 shewanella
34	46.5	62.0	252	2	Q88DQ3_PSEPK	Q88dq3 pseudomonas
35	46	61.3	157	2	Q74XC9_YERPE	Q74xc9 yersinia pe
36	46	61.3	165	2	Q40425_NICSY	Q40425 nicotiana s
37	46	61.3	181	2	Q8D1E6_YERPE	Q8d1e6 yersinia pe
38	46	61.3	192	2	Q8ZIY5_YERPE	Q8ziy5 yersinia pe
39	46	61.3	192	2	Q66FD7_YERPS	Q66fd7 yersinia ps
40	46	61.3	287	2	Q17200_BOMMO	Q17200 bombyx mori
41	46	61.3	296	2	Q6AVS5_ORYSA	Q6avs5 oryza sativ
42	46	61.3	303	2	Q17201_BOMMO	Q17201 bombyx mori
43	46	61.3	315	2	Q8IRS1_DROME	Q8irs1 drosophila
44	46	61.3	347	2	Q9GZC7_TRYCR	Q9gzc7 trypanosoma
45	46	61.3	396	2	Q8DIF7_SYNEL	Q8dif7 synechococc
46	46	61.3	460	2	Q55MV2_CRYNE	Q55mv2 cryptococcu
47	46	61.3	460	2	Q5KB79_CRYNE	Q5kb79 cryptococcu
48	46	61.3	524	2	Q8EH30_SHEON	Q8eh30 shewanella
49	46	61.3	656	2	Q7R3F3_GIALA	Q7r3f3 giardia lam
50	46	61.3	945	2	Q8X087_NEUCR	Q8x087 neurospora
51	46	61.3	1048	2	Q9VX90_DROME	Q9vx90 drosophila
52	46	61.3	1077	2	Q8IR04_DROME	Q8ir04 drosophila
53	46	61.3	1701	2	Q7SCH8_NEUCR	Q7sch8 neurospora
54	46	61.3	1742	2	Q5AVF0_EMENI	Q5avf0 aspergillus
55	45.5	60.7	138	1	FLAV_CLOBE	P00322 clostridium
56	45.5	60.7	263	2	Q6ZL79_ORYSA	Q6zl79 oryza sativ
57	45	60.0	71	2	Q612A0_CAEBR	Q612a0 caenorhabdi
58	45	60.0	71	2	Q18838_CAEEL	Q18838 caenorhabdi
59	45	60.0	185	2	Q9SIX3_ARATH	Q9six3 arabidopsis
60	45	60.0	194	2	O96853_SCHHA	O96853 schistosoma
61	45	60.0	240	2	Q4WLC1_ASPFU	Q4wlc1 aspergillus
62	45	60.0	259	2	Q7QCR4_ANOGA	Q7qcr4 anopheles g
63	45	60.0	268	2	Q51K94_MAGGR	Q51k94 magnaporthe
64	45	60.0	278	2	Q7QCR3_ANOGA	Q7qcr3 anopheles g
65	45	60.0	289	2	Q9C909_ARATH	Q9c909 arabidopsis
66	45	60.0	304	2	Q7V9Z7_PROMA	Q7v9z7 prochloroco
67	45	60.0	309	2	Q9FNR1_ARATH	Q9fnr1 arabidopsis
68	45	60.0	331	2	Q8KHE9_SHIFL	Q8khe9 shigella fl
69	45	60.0	331	2	Q8KHF0_SHIDY	Q8khf0 shigella dy
70	45	60.0	331	2	Q8KXF9_ECOLI	Q8kxf9 escherichia
71	45	60.0	331	2	Q8KXG1_ECOLI	Q8kxg1 escherichia
72	45	60.0	331	2	Q8KXG2_ECOLI	Q8kxg2 escherichia

```
 73      45    60.0    331    2    Q8KXG3_ECOLI          Q8kxg3 escherichia
 74      45    60.0    331    2    Q8KXG5_SHISO          Q8kxg5 shigella so
 75      45    60.0    331    2    Q8KXG6_SHISO          Q8kxg6 shigella so
 76      45    60.0    331    2    Q8KXG7_SHISO          Q8kxg7 shigella so
 77      45    60.0    331    2    Q8KXG9_SHISO          Q8kxg9 shigella so
 78      45    60.0    331    2    Q8KXH0_SHISO          Q8kxh0 shigella so
 79      45    60.0    331    2    Q8KXH1_SHIDY          Q8kxh1 shigella dy
 80      45    60.0    331    2    Q8KXH2_SHIDY          Q8kxh2 shigella dy
 81      45    60.0    331    2    Q8KXH3_SHIDY          Q8kxh3 shigella dy
 82      45    60.0    331    2    Q8KXH5_SHIFL          Q8kxh5 shigella fl
 83      45    60.0    331    2    Q8KXH6_SHIFL          Q8kxh6 shigella fl
 84      45    60.0    331    2    Q8KXH7_SHIFL          Q8kxh7 shigella fl
 85      45    60.0    331    2    Q8KXH8_SHIFL          Q8kxh8 shigella fl
 86      45    60.0    331    2    Q8KXH9_SHIFL          Q8kxh9 shigella fl
 87      45    60.0    331    2    Q8KXI1_SHIFL          Q8kxi1 shigella fl
 88      45    60.0    331    2    Q8KXI2_SHIBO          Q8kxi2 shigella bo
 89      45    60.0    331    2    Q8KXI3_SHIBO          Q8kxi3 shigella bo
 90      45    60.0    331    2    Q8KXI4_SHIBO          Q8kxi4 shigella bo
 91      45    60.0    331    2    Q8KXI5_SHIBO          Q8kxi5 shigella bo
 92      45    60.0    331    2    Q8KXI6_SHIBO          Q8kxi6 shigella bo
 93      45    60.0    331    2    Q8KXI7_SHIBO          Q8kxi7 shigella bo
 94      45    60.0    331    2    Q8KHF1_SHIBO          Q8khf1 shigella bo
 95      45    60.0    386    2    Q7PZ31_ANOGA          Q7pz31 anopheles g
 96      45    60.0    408    2    Q8TVT4_METKA          Q8tvt4 methanopyru
 97      45    60.0    410    2    Q52D09_MAGGR          Q52d09 magnaporthe
 98      45    60.0    422    2    Q4JIR9_9BACT          Q4jir9 uncultured
 99      45    60.0    498    2    Q7VI48_HELHP          Q7vi48 helicobacte
100      45    60.0    499    2    Q8DXY8_STRA5          Q8dxy8 streptococc
```

ALIGNMENTS

RESULT 1
Q93119_ANTPE
ID   Q93119_ANTPE PRELIMINARY;   PRT;   421 AA.
AC   Q93119;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Antheraea pernyi fibroin (Fragment).
OS   Antheraea pernyi (Chinese oak silk moth).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea;
OC   Saturniidae; Saturniinae; Saturniini; Antheraea.
OX   NCBI_TaxID=7119;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Posterior silkglands;
RX   MEDLINE=97165499; PubMed=9013260;
RA   Yukuhiro K., Kanda T., Tamura T.;
RT   "Preferential codon usage and two types of repetitive motifs in the
RT   fibroin gene of the Chinese oak silkworm, Antheraea pernyi.";
RL   Insect Mol. Biol. 6:89-95(1997).
DR   EMBL; D83241; BAA11860.1; -; mRNA.
DR   HSSP; O87916; 1JTD.
FT   NON_TER   1   1

```
SQ    SEQUENCE    421 AA;   35800 MW;   6FBA092830262D8E CRC64;

   Query Match                 100.0%;   Score 75;   DB 2;   Length 421;
   Best Local Similarity   100.0%;   Pred. No. 0.0061;
   Matches   12;  Conservative    0;   Mismatches    0;   Indels    0;  Gaps    0;

Qy        1 YGWGDGGYGSDS 12
            ||||||||||||
Db       48 YGWGDGGYGSDS 59


Search completed: December  8, 2005, 08:15:39
Job time : 163.273 secs
```

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:04:51 ; Search time 25.6364 Seconds
                                              (without alignments)
                                              45.038 Million cell updates/sec

Title: US-10-789-494B-5
Perfect score: 75
Sequence: 1 YGWGDGGYGSDS 12

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   PIR_80:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	75	100.0	2639	2	T31328	fibroin - Chinese
2	52.5	70.0	142	2	B96972	flavodoxin [import
3	51	68.0	169	1	S38331	glycine-rich RNA-b
4	50	66.7	602	2	T35782	probable secreted
5	48	64.0	214	1	KNNT2S	glycine-rich prote
6	46	61.3	165	2	S59529	RNA-binding glycin
7	46	61.3	165	2	S41773	glycine-rich RNA-b
8	46	61.3	192	2	AE0043	probable membrane
9	45.5	60.7	138	1	FXCLEX	flavodoxin - Clost
10	45	60.0	71	2	T15836	hypothetical prote
11	45	60.0	185	2	D84538	probable glycine-r
12	45	60.0	220	2	A44805	eggshell protein p
13	45	60.0	289	2	F96770	protein RNA-bindin

14	45	60.0	1102	2	A32247	virG protein - Shi
15	44.5	59.3	1108	2	D96798	hypothetical prote
16	44	58.7	142	2	S12311	glycine-rich RNA-b
17	44	58.7	327	2	T04919	DNA-binding protei
18	44	58.7	345	2	T07839	ananain (EC 3.4.22
19	44	58.7	407	2	F85079	probable transposo
20	44	58.7	434	2	E70768	hypothetical glyci
21	44	58.7	440	2	T50662	UVB-resistance pro
22	44	58.7	751	2	F87789	protein C34G6.2 [i
23	44	58.7	4861	2	S71752	giant protein p619
24	43.5	58.0	1028	2	A96719	hypothetical prote
25	43.5	58.0	1433	2	A46053	bullous pemphigoid
26	43	57.3	160	2	F64816	ybiA protein - Esc
27	43	57.3	447	2	T00435	probable mitochond
28	43	57.3	502	2	A70582	hypothetical prote
29	43	57.3	527	2	B70700	hypothetical prote
30	43	57.3	614	2	T10862	phaseolin G-box bi
31	43	57.3	1055	2	S53597	chlorophyll a/b-bi
32	43	57.3	4836	2	T14346	herc2 protein - mo
33	42.5	56.7	1684	2	T02367	hypothetical prote
34	42	56.0	63	2	S44634	f22b7.4 protein -
35	42	56.0	150	2	T03586	glycine-rich RNA-b
36	42	56.0	193	2	S24295	chorion protein -
37	42	56.0	211	2	S21864	probable cysteine
38	42	56.0	245	2	JQ0337	allergen Der p 1 -
39	42	56.0	257	2	C84533	hypothetical prote
40	42	56.0	285	2	T31503	hypothetical prote
41	42	56.0	313	2	S47433	cathepsin L (EC 3.
42	42	56.0	319	2	A61500	allergen Der f I p
43	42	56.0	333	2	T50630	hypothetical prote
44	42	56.0	350	2	T16385	hypothetical prote
45	42	56.0	392	2	G95258	secreted 45 kd pro
46	42	56.0	392	2	B98124	general stress pro
47	42	56.0	395	2	H84765	hypothetical prote
48	42	56.0	463	2	T46290	hypothetical prote
49	42	56.0	730	2	T43317	pgl-1 protein - Ca
50	42	56.0	771	2	T29177	hypothetical prote
51	42	56.0	968	2	E90481	alpha-mannosidase
52	42	56.0	1032	2	AI1697	alpha-mannosidase
53	42	56.0	1036	2	AG1326	alpha-mannosidase
54	42	56.0	1039	2	G83748	alpha-mannosidase
55	42	56.0	1660	2	A70869	hypothetical glyci
56	41.5	55.3	23	2	A32473	histidine-rich pro
57	41	54.7	49	2	T02026	glycine-rich prote
58	41	54.7	82	2	S19774	glycine-rich prote
59	41	54.7	139	2	S31443	glycine-rich RNA-b
60	41	54.7	144	2	S77128	hypothetical prote
61	41	54.7	145	2	T01356	glycine-rich RNA b
62	41	54.7	148	2	S41772	glycine-rich RNA-b
63	41	54.7	154	2	E84468	probable glycine-r
64	41	54.7	155	2	S20846	glycine-rich prote
65	41	54.7	157	1	S14857	glycine-rich prote
66	41	54.7	158	2	T05254	probable RNA-bindi
67	41	54.7	161	2	S71453	glycine-rich RNA-b
68	41	54.7	166	2	T10463	glycine-rich prote
69	41	54.7	167	2	S71779	glycine-rich RNA-b
70	41	54.7	168	1	S12312	glycine-rich RNA-b

```
 71     41    54.7    169    2    T10465              glycine-rich prote
 72     41    54.7    173    2    S53050              RNA binding protei
 73     41    54.7    175    2    S54255              probable glycine r
 74     41    54.7    179    2    T05810              hypothetical prote
 75     41    54.7    203    1    JQ1061              glycine-rich prote
 76     41    54.7    259    2    T15126              hypothetical prote
 77     41    54.7    281    2    A65219              phnJ protein - Esc
 78     41    54.7    281    2    A91264              phosphonate metabo
 79     41    54.7    281    2    F86104              phosphonate metabo
 80     41    54.7    288    2    AE2083              phosphonate metabo
 81     41    54.7    293    2    AE0420              PhnJ protein [impo
 82     41    54.7    294    2    C83224              conserved hypothet
 83     41    54.7    297    2    F96023              probable C-P (carb
 84     41    54.7    300    2    AD2598              conserved hypothet
 85     41    54.7    300    2    E97380              phnJ protein [impo
 86     41    54.7    305    2    T06413              cathepsin B-like c
 87     41    54.7    307    2    A32208              synaptophysin - bo
 88     41    54.7    307    2    B27287              synaptophysin - ra
 89     41    54.7    312    2    AC2228              hypothetical prote
 90     41    54.7    313    2    A35699              synaptophysin - hu
 91     41    54.7    315    2    I50706              transcription fact
 92     41    54.7    346    1    S35500              heterogeneous ribo
 93     41    54.7    353    2    T06466              cathepsin B-like c
 94     41    54.7    356    2    S60479              cathepsin B-like c
 95     41    54.7    359    2    T02011              probable cathepsin
 96     41    54.7    382    2    T27076              hypothetical prote
 97     41    54.7    407    2    T21956              hypothetical prote
 98     41    54.7    444    2    C90704              Rhs core protein [
 99     41    54.7    444    2    G85554              Rhs core protein [
100     41    54.7    532    2    S46954              transposase - Caen
```

ALIGNMENTS

RESULT 1
T31328
fibroin - Chinese oak silkmoth
C;Species: Antheraea pernyi (Chinese oak silkmoth)
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T31328
R;Sezutsu, H.; Tamura, T.; Yukuhiro, K.
submitted to the EMBL Data Library, August 1998
A;Description: Characterization of the full length fibroin gene of a wild
silkworm, Antheraea pernyi.
A;Reference number: Z20995
A;Accession: T31328
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2639 <SEZ>
A;Cross-references: UNIPROT:O76786; UNIPARC:UPI0000078D8E; EMBL:AF083334;
NID:g3450882; PID:g3450883; PIDN:AAC32606.1
C;Genetics:
A;Introns: 14/3

    Query Match   100.0%; Score 75; DB 2; Length 2639;
    Best Local Similarity 100.0%; Pred. No. 0.011;

```
   Matches    12;  Conservative     0;  Mismatches     0;  Indels     0;  Gaps     0;

Qy              1 YGWGDGGYGSDS 12
                  ||||||||||||
Db            257 YGWGDGGYGSDS 268


RESULT 2
B96972
flavodoxin [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: B96972
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson,
R.; Lee, H.M.; Dubois, J.; Qiu, D.; Hitti, J.; Wolf, Y.I.; Tatusov, R.L.;
Sabathe, F.; Doucette-Stamm, L.; Soucaille, P.; Daly, M.J.; Bennett, G.N.;
Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing
Bacterium Clostridium acetobutylicum.
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: B96972
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-142 <KUR>
A;Cross-references: UNIPROT:Q97LH4; UNIPARC:UPI00000C9EF3; GB:AE001437;
PIDN:AAK78565.1; PID:g15023456; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0587
C;Superfamily: flavodoxin; flavodoxin homology
C;Keywords: flavoprotein

   Query Match            70.0%;  Score 52.5;  DB 2;  Length 142;
   Best Local Similarity  45.5%;  Pred. No. 1.2;
   Matches    10;  Conservative     1;  Mismatches     0;  Indels    11;  Gaps     1;

Qy              1 YGWGDG-----------GYGSD 11
                  ||||||           |||:|
Db             91 YGWGDGQFMRDWVERMEGYGAD 112


Search completed: December  8, 2005, 08:16:30
Job time : 28.6364 secs
```

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:16:37 ; Search time 7.09091 Seconds
                                              (without alignments)
                                              9.451 Million cell updates/sec

Title: US-10-789-494B-5
Perfect score: 75
Sequence: 1 YGWGDGGYGSDS 12

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 32527 seqs, 5584426 residues

Total number of hits satisfying chosen parameters: 32527

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   . Published_Applications_AA_New:*
                1: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                4: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                5: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
                8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                              SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	42	56.0	576	7	US-11-137-465-64	Sequence 64, Appl
2	42	56.0	579	7	US-11-045-802-32	Sequence 32, Appl
3	42	56.0	579	7	US-11-045-802-33	Sequence 33, Appl
4	42	56.0	634	7	US-11-137-465-63	Sequence 63, Appl
5	40	53.3	27	6	US-10-939-890-352	Sequence 352, App
6	40	53.3	27	6	US-10-939-890-828	Sequence 828, App
7	40	53.3	397	6	US-10-467-657-3814	Sequence 3814, Ap
8	39	52.0	123	6	US-10-467-657-8536	Sequence 8536, Ap
9	39	52.0	411	7	US-11-092-140-114	Sequence 114, App

10	39	52.0	488	7	US-11-092-140-110	Sequence 110, App
11	38.5	51.3	488	6	US-10-821-234-1000	Sequence 1000, Ap
12	38	50.7	37	7	US-11-110-424-16	Sequence 16, Appl
13	38	50.7	42	7	US-11-110-424-15	Sequence 15, Appl
14	38	50.7	2725	7	US-11-113-424-52	Sequence 52, Appl
15	37.5	50.0	390	6	US-10-821-234-1053	Sequence 1053, Ap
16	37	49.3	11	7	US-11-035-546-18	Sequence 18, Appl
17	37	49.3	229	6	US-10-689-742-38	Sequence 38, Appl
18	37	49.3	262	6	US-10-689-742-65	Sequence 65, Appl
19	37	49.3	522	7	US-11-080-991-104	Sequence 104, App
20	37	49.3	658	6	US-10-467-657-4160	Sequence 4160, Ap
21	37	49.3	660	6	US-10-878-556A-102	Sequence 102, App
22	37	49.3	674	6	US-10-821-234-965	Sequence 965, App
23	36	48.0	346	6	US-10-878-556A-58	Sequence 58, Appl
24	36	48.0	486	6	US-10-821-234-1063	Sequence 1063, Ap
25	36	48.0	544	6	US-10-821-234-889	Sequence 889, App
26	36	48.0	760	7	US-11-186-284-55	Sequence 55, Appl
27	36	48.0	2516	6	US-10-647-956A-2	Sequence 2, Appli
28	35	46.7	41	7	US-11-098-674-17	Sequence 17, Appl
29	35	46.7	120	6	US-10-821-234-1243	Sequence 1243, Ap
30	35	46.7	211	7	US-11-089-551A-19	Sequence 19, Appl
31	35	46.7	212	7	US-11-089-551A-21	Sequence 21, Appl
32	35	46.7	237	7	US-11-054-515-1906	Sequence 1906, Ap
33	35	46.7	237	7	US-11-054-515-2003	Sequence 2003, Ap
34	35	46.7	237	7	US-11-054-515-2005	Sequence 2005, Ap
35	35	46.7	237	7	US-11-054-515-2017	Sequence 2017, Ap
36	35	46.7	237	7	US-11-054-515-2019	Sequence 2019, Ap
37	35	46.7	237	7	US-11-054-515-2020	Sequence 2020, Ap
38	35	46.7	237	7	US-11-054-515-2027	Sequence 2027, Ap
39	35	46.7	237	7	US-11-054-515-2028	Sequence 2028, Ap
40	35	46.7	237	7	US-11-054-515-2040	Sequence 2040, Ap
41	35	46.7	237	7	US-11-054-515-2043	Sequence 2043, Ap
42	35	46.7	237	7	US-11-054-515-2104	Sequence 2104, Ap
43	35	46.7	237	7	US-11-054-515-2110	Sequence 2110, Ap
44	35	46.7	237	7	US-11-054-515-2111	Sequence 2111, Ap
45	35	46.7	237	7	US-11-054-515-2114	Sequence 2114, Ap
46	35	46.7	237	7	US-11-054-515-2115	Sequence 2115, Ap
47	35	46.7	237	7	US-11-054-515-2118	Sequence 2118, Ap
48	35	46.7	238	7	US-11-054-515-2053	Sequence 2053, Ap
49	35	46.7	239	7	US-11-054-515-1882	Sequence 1882, Ap
50	35	46.7	239	7	US-11-054-515-1922	Sequence 1922, Ap
51	35	46.7	241	7	US-11-054-515-1577	Sequence 1577, Ap
52	35	46.7	241	7	US-11-054-515-1887	Sequence 1887, Ap
53	35	46.7	241	7	US-11-054-515-1889	Sequence 1889, Ap
54	35	46.7	241	7	US-11-054-515-1901	Sequence 1901, Ap
55	35	46.7	241	7	US-11-054-515-1932	Sequence 1932, Ap
56	35	46.7	241	7	US-11-054-515-1948	Sequence 1948, Ap
57	35	46.7	241	7	US-11-054-515-2054	Sequence 2054, Ap
58	35	46.7	242	7	US-11-054-515-1580	Sequence 1580, Ap
59	35	46.7	242	7	US-11-054-515-1915	Sequence 1915, Ap
60	35	46.7	243	7	US-11-054-515-1883	Sequence 1883, Ap
61	35	46.7	243	7	US-11-054-515-1935	Sequence 1935, Ap
62	35	46.7	243	7	US-11-054-515-1943	Sequence 1943, Ap
63	35	46.7	243	7	US-11-054-515-1945	Sequence 1945, Ap
64	35	46.7	243	7	US-11-054-515-2063	Sequence 2063, Ap
65	35	46.7	243	7	US-11-054-515-2064	Sequence 2064, Ap
66	35	46.7	244	7	US-11-054-515-82	Sequence 82, Appl

```
 67      35     46.7     244    7   US-11-054-515-164      Sequence 164, App
 68      35     46.7     244    7   US-11-054-515-261      Sequence 261, App
 69      35     46.7     244    7   US-11-054-515-280      Sequence 280, App
 70      35     46.7     244    7   US-11-054-515-1881     Sequence 1881, Ap
 71      35     46.7     245    7   US-11-054-515-1291     Sequence 1291, Ap
 72      35     46.7     245    7   US-11-054-515-1900     Sequence 1900, Ap
 73      35     46.7     245    7   US-11-054-515-1902     Sequence 1902, Ap
 74      35     46.7     246    7   US-11-054-515-1920     Sequence 1920, Ap
 75      35     46.7     246    7   US-11-054-515-2062     Sequence 2062, Ap
 76      35     46.7     247    7   US-11-054-515-969      Sequence 969, App
 77      35     46.7     247    7   US-11-054-515-1177     Sequence 1177, Ap
 78      35     46.7     247    7   US-11-054-515-1923     Sequence 1923, Ap
 79      35     46.7     247    7   US-11-054-515-3240     Sequence 3240, Ap
 80      35     46.7     247    7   US-11-054-515-3242     Sequence 3242, Ap
 81      35     46.7     248    6   US-10-512-184-32       Sequence 32, Appl
 82      35     46.7     248    7   US-11-054-515-1004     Sequence 1004, Ap
 83      35     46.7     248    7   US-11-054-515-1008     Sequence 1008, Ap
 84      35     46.7     248    7   US-11-054-515-1104     Sequence 1104, Ap
 85      35     46.7     248    7   US-11-054-515-1385     Sequence 1385, Ap
 86      35     46.7     248    7   US-11-054-515-1421     Sequence 1421, Ap
 87      35     46.7     248    7   US-11-054-515-1446     Sequence 1446, Ap
 88      35     46.7     248    7   US-11-054-515-1700     Sequence 1700, Ap
 89      35     46.7     248    7   US-11-054-515-1717     Sequence 1717, Ap
 90      35     46.7     248    7   US-11-054-515-1719     Sequence 1719, Ap
 91      35     46.7     248    7   US-11-054-515-1733     Sequence 1733, Ap
 92      35     46.7     248    7   US-11-054-515-1737     Sequence 1737, Ap
 93      35     46.7     248    7   US-11-054-515-1771     Sequence 1771, Ap
 94      35     46.7     248    7   US-11-054-515-1778     Sequence 1778, Ap
 95      35     46.7     248    7   US-11-054-515-1876     Sequence 1876, Ap
 96      35     46.7     248    7   US-11-054-515-1879     Sequence 1879, Ap
 97      35     46.7     249    7   US-11-054-515-414      Sequence 414, App
 98      35     46.7     249    7   US-11-054-515-918      Sequence 918, App
 99      35     46.7     249    7   US-11-054-515-919      Sequence 919, App
100      35     46.7     249    7   US-11-054-515-926      Sequence 926, App
```

ALIGNMENTS

RESULT 1
US-11-137-465-64
; Sequence 64, Application US/11137465
; Publication No. US20050255558A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdoch, Paul R.
; APPLICANT: Rizvi, Safia, K.
; APPLICANT: Smith, Randall, F.
; APPLICANT: Xiang, Zhaoying
; APPLICANT: Kabnick, Karen
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50018
; CURRENT APPLICATION NUMBER: US/11/137,465
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US/10/239,663
; PRIOR FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/US01/09226

    Query Match   56.0%; Score 42; DB 7; Length 576;
    Best Local Similarity   64.3%; Pred. No. 20;
    Matches   9; Conservative   0; Mismatches   1; Indels   4; Gaps   1;

Qy   2 GWGDGGY----GSD 11
              |||| || |||
Db   441 GWGDIGYSFVVGSD 454

RESULT 5
US-10-939-890-352
; Sequence 352, Application US/10939890
; Publication No. US20050250700A1
; GENERAL INFORMATION:
; APPLICANT: Sato, Aaron K.
; APPLICANT: Sexton, Daniel J.
; APPLICANT: Dransfield, Daniel T.
; APPLICANT: Ladner, Robert C.
; APPLICANT: Arbogast, Christophe
; APPLICANT: Bussat, Philippe
; APPLICANT: Fan, Hong
; APPLICANT: Khurana, Sudha
; APPLICANT: Linder, Karen E.
; APPLICANT: Marinelli, Edmund R.
; APPLICANT: Nanjappan, Palaniappa
; APPLICANT: Nunn, Adrian D.
; APPLICANT: Pillai, Radhakrishna
; APPLICANT: Pochon, Sibylle
; APPLICANT: Ramalingam, Kondareddiar
; APPLICANT: Shrivastava, Ajay
; APPLICANT: Song, Bo
; APPLICANT: Swenson, Rolf E.
; APPLICANT: Von Wronski, Mathew A.
; TITLE OF INVENTION: KDR AND VEGF/KDR BINDING PEPTIDES
; FILE REFERENCE: D0617.70014US00
; CURRENT APPLICATION NUMBER: US/10/939,890
; CURRENT FILING DATE: 2004-09-13
; PRIOR APPLICATION NUMBER: US 10/661,156
; PRIOR FILING DATE: 2003-09-11
; PRIOR APPLICATION NUMBER: US 10/382,082
; PRIOR FILING DATE: 2003-03-03

```
  Query Match               53.3%;  Score 40;  DB 6;  Length 27;
  Best Local Similarity    66.7%;  Pred. No. 2.9;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;


Qy          1 YGWGDGGYG 9
              :|| ||| |
Db         17 WGWADGGGG 25
```

RESULT 12
US-11-110-424-16

```
  Query Match               50.7%;  Score 38;  DB 7;  Length 37;
```

```
Best Local Similarity    75.0%;  Pred. No. 6.9;
  Matches     6;  Conservative     1;  Mismatches     1;  Indels     0;  Gaps     0;

Qy          2 GWGDGGYG 9
              |:| ||||
Db          9 GYGPGGYG 16

Search completed: December  8, 2005, 08:35:42
Job time : 8.09091 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:15:47 ; Search time 121.636 Seconds
                                             (without alignments)
                                             41.221 Million cell updates/sec

Title: US-10-789-494B-5
Perfect score: 75
Sequence: 1 YGWGDGGYGSDS 12

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   Published_Applications_AA_Main:*
                 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
                 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
                 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
                 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                              SUMMARIES
                    %
Result   Query
   No.   Score   Match Length DB   ID   Description
----
      1   75   100.0   12 5   US-10-789-494B-5   Sequence 5, Appli
      2   75   100.0   12 5   US-10-789-494B-62   Sequence 62, Appl
      3   75   100.0   14 5   US-10-789-494B-32   Sequence 32, Appl
      4   75   100.0   21 5   US-10-789-494B-30   Sequence 30, Appl
      5   75   100.0   21 5   US-10-789-494B-34   Sequence 34, Appl
      6   75   100.0   21 5   US-10-789-494B-37   Sequence 37, Appl
      7   75   100.0   21 5   US-10-789-494B-44   Sequence 44, Appl
      8   75   100.0   22 5   US-10-789-494B-36   Sequence 36, Appl
      9   75   100.0   22 5   US-10-789-494B-40   Sequence 40, Appl
     10   75   100.0   22 5   US-10-789-494B-42   Sequence 42, Appl
     11   75   100.0   22 5   US-10-789-494B-53   Sequence 53, Appl

12	75	100.0	45	5	US-10-789-494B-25	Sequence 25, Appl
13	71	94.7	21	5	US-10-789-494B-35	Sequence 35, Appl
14	70	93.3	21	5	US-10-789-494B-45	Sequence 45, Appl
15	69	92.0	21	5	US-10-789-494B-46	Sequence 46, Appl
16	68	90.7	21	5	US-10-789-494B-39	Sequence 39, Appl
17	63.5	84.7	25	5	US-10-789-494B-48	Sequence 48, Appl
18	59	78.7	22	5	US-10-789-494B-52	Sequence 52, Appl
19	55	73.3	23	5	US-10-789-494B-28	Sequence 28, Appl
20	53	70.7	21	5	US-10-789-494B-51	Sequence 51, Appl
21	52	69.3	114	4	US-10-437-963-174027	Sequence 174027,
22	51	68.0	170	5	US-10-739-930-6860	Sequence 6860, Ap
23	50	66.7	599	4	US-10-156-761-9345	Sequence 9345, Ap
24	48	64.0	23	5	US-10-789-494B-49	Sequence 49, Appl
25	48	64.0	85	4	US-10-424-599-209214	Sequence 209214,
26	48	64.0	129	4	US-10-425-114-69985	Sequence 69985, A
27	48	64.0	163	4	US-10-767-701-45098	Sequence 45098, A
28	48	64.0	472	4	US-10-425-115-316355	Sequence 316355,
29	48	64.0	493	4	US-10-425-114-68940	Sequence 68940, A
30	48	64.0	494	4	US-10-425-114-68603	Sequence 68603, A
31	47	62.7	16	5	US-10-789-494B-41	Sequence 41, Appl
32	47	62.7	17	5	US-10-789-494B-26	Sequence 26, Appl
33	47	62.7	17	5	US-10-789-494B-64	Sequence 64, Appl
34	47	62.7	183	4	US-10-425-115-289352	Sequence 289352,
35	47	62.7	462	4	US-10-424-599-207928	Sequence 207928,
36	47	62.7	806	4	US-10-220-480-42	Sequence 42, Appl
37	47	62.7	806	4	US-10-220-481-147	Sequence 147, App
38	47	62.7	998	4	US-10-369-493-13474	Sequence 13474, A
39	47	62.7	1022	4	US-10-156-761-9292	Sequence 9292, Ap
40	46	61.3	238	5	US-10-476-264-112	Sequence 112, App
41	46	61.3	256	4	US-10-437-963-194431	Sequence 194431,
42	46	61.3	671	4	US-10-437-963-109378	Sequence 109378,
43	46	61.3	1048	6	US-11-097-143-19395	Sequence 19395, A
44	45	60.0	100	4	US-10-767-701-42596	Sequence 42596, A
45	45	60.0	133	4	US-10-424-599-281181	Sequence 281181,
46	45	60.0	168	4	US-10-425-115-264232	Sequence 264232,
47	45	60.0	1004	4	US-10-156-761-14806	Sequence 14806, A
48	44.5	59.3	157	4	US-10-424-599-158770	Sequence 158770,
49	44.5	59.3	636	4	US-10-424-599-225665	Sequence 225665,
50	44.5	59.3	1103	5	US-10-739-930-6660	Sequence 6660, Ap
51	44	58.7	91	4	US-10-424-599-162821	Sequence 162821,
52	44	58.7	120	4	US-10-425-115-268850	Sequence 268850,
53	44	58.7	130	4	US-10-424-599-221597	Sequence 221597,
54	44	58.7	141	4	US-10-437-963-120280	Sequence 120280,
55	44	58.7	141	4	US-10-425-115-243677	Sequence 243677,
56	44	58.7	141	6	US-11-097-143-39207	Sequence 39207, A
57	44	58.7	144	4	US-10-437-963-124401	Sequence 124401,
58	44	58.7	157	4	US-10-767-701-32016	Sequence 32016, A
59	44	58.7	166	4	US-10-767-701-47009	Sequence 47009, A
60	44	58.7	248	4	US-10-425-115-274617	Sequence 274617,
61	44	58.7	298	4	US-10-424-599-275592	Sequence 275592,
62	44	58.7	302	4	US-10-425-114-67880	Sequence 67880, A
63	44	58.7	318	3	US-09-934-455-362	Sequence 362, App
64	44	58.7	318	4	US-10-278-173-42	Sequence 42, Appl
65	44	58.7	318	4	US-10-225-066A-390	Sequence 390, App
66	44	58.7	318	4	US-10-374-780A-2250	Sequence 2250, Ap
67	44	58.7	318	4	US-10-412-699B-96	Sequence 96, Appl
68	44	58.7	318	5	US-10-495-918-130	Sequence 130, App

```
69     44    58.7     318    5    US-10-225-066A-390       Sequence 390, App
70     44    58.7     330    4    US-10-425-114-60210      Sequence 60210, A
71     44    58.7     369    4    US-10-437-963-191722     Sequence 191722,
72     44    58.7     441    4    US-10-424-599-216853     Sequence 216853,
73     44    58.7     443    4    US-10-425-115-287505     Sequence 287505,
74     44    58.7     444    5    US-10-732-923-10430      Sequence 10430, A
75     44    58.7     453    4    US-10-437-963-125860     Sequence 125860,
76     44    58.7     453    4    US-10-437-963-202405     Sequence 202405,
77     44    58.7     457    4    US-10-425-115-312101     Sequence 312101,
78     44    58.7     459    4    US-10-425-114-47823      Sequence 47823, A
79     44    58.7     479    4    US-10-424-599-188179     Sequence 188179,
80     44    58.7     492    4    US-10-425-114-69226      Sequence 69226, A
81     44    58.7     521    4    US-10-437-963-117262     Sequence 117262,
82     44    58.7     523    4    US-10-437-963-116479     Sequence 116479,
83     44    58.7     530    4    US-10-425-114-65678      Sequence 65678, A
84     44    58.7     534    5    US-10-739-930-10894      Sequence 10894, A
85     44    58.7     548    4    US-10-425-115-274613     Sequence 274613,
86     44    58.7     565    4    US-10-424-599-275593     Sequence 275593,
87     44    58.7     634    4    US-10-094-749-2263       Sequence 2263, Ap
88     44    58.7     776    4    US-10-424-599-180732     Sequence 180732,
89     44    58.7    1170    4    US-10-369-493-3852       Sequence 3852, Ap
90     44    58.7    1357    4    US-10-424-599-186063     Sequence 186063,
91     44    58.7    4861    3    US-09-919-497-70         Sequence 70, Appl
92     44    58.7    4861    4    US-10-097-534-26         Sequence 26, Appl
93     44    58.7    4861    4    US-10-146-473-49         Sequence 49, Appl
94     44    58.7    4861    5    US-10-287-436A-486       Sequence 486, App
95     44    58.7    4861    5    US-10-287-436A-1182      Sequence 1182, Ap
96     44    58.7    4899    6    US-11-097-143-24447      Sequence 24447, A
97   43.5    58.0     137    4    US-10-767-701-55069      Sequence 55069, A
98   43.5    58.0     218    4    US-10-425-115-347280     Sequence 347280,
99   43.5    58.0     288    4    US-10-425-115-347272     Sequence 347272,
100  43.5    58.0     354    4    US-10-425-115-347277     Sequence 347277,
```

ALIGNMENTS

RESULT 1
US-10-789-494B-5
; Sequence 5, Application US/10789494B
; Publication No. US20050143296A1
; GENERAL INFORMATION:
; APPLICANT: TSUBOUCHI, Kozo
; APPLICANT: YAMADA, Hiromi
; TITLE OF INVENTION: EXTRACTION AND UTILIZATION OF CELL
; TITLE OF INVENTION: GROWTH-PROMOTING PEPTIDES FROM SILK PROTEIN
; FILE REFERENCE: OPS 635
; CURRENT APPLICATION NUMBER: US/10/789,494B
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: JP 2003-55048
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 5
;   LENGTH: 12
;   TYPE: PRT
;   ORGANISM: Antheraea yamamai
US-10-789-494B-5

```
Query Match              100.0%;  Score 75;  DB 5;  Length 12;
Best Local Similarity    100.0%;  Pred. No. 0.0013;
Matches   12;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;


Qy        1 YGWGDGGYGSDS 12
            ||||||||||||
Db        1 YGWGDGGYGSDS 12




RESULT 21
US-10-437-963-174027
; Sequence 174027, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 174027
;   LENGTH: 114
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_72008C.1.pep
US-10-437-963-174027

  Query Match              69.3%;  Score 52;  DB 4;  Length 114;
  Best Local Similarity    88.9%;  Pred. No. 13;
  Matches    8;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy        1 YGWGDGGYG 9
            |||||  |||
Db       78 YGWGDCGYG 86


Search completed: December  8, 2005, 08:35:25
Job time : 122.636 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:07:46 ; Search time 36 Seconds
                                              (without alignments)
                                              27.559 Million cell updates/sec

Title: US-10-789-494B-5
Perfect score: 75
Sequence: 1 YGWGDGGYGSDS 12

Scoring table: BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   Issued_Patents_AA:*
            1:    /cgn2_6/ptodata/1/iaa/5_COMB.pep:*
            2:    /cgn2_6/ptodata/1/iaa/6_COMB.pep:*
            3:    /cgn2_6/ptodata/1/iaa/H_COMB.pep:*
            4:    /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
            5:    /cgn2_6/ptodata/1/iaa/RE_COMB.pep:*
            6:    /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                  SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	46	61.3	158	2	US-09-270-767-33522	Sequence 33522, A
2	44	58.7	4861	2	US-09-919-497-70	Sequence 70, Appl
3	43	57.3	135	2	US-09-270-767-32080	Sequence 32080, A
4	43	57.3	286	2	US-09-640-211A-920	Sequence 920, App
5	43	57.3	527	2	US-09-712-363-156	Sequence 156, App
6	42.5	56.7	274	2	US-09-976-594-417	Sequence 417, App
7	42.5	56.7	378	2	US-10-164-595-2	Sequence 2, Appli
8	42	56.0	24	1	US-08-482-142-25	Sequence 25, Appl
9	42	56.0	24	1	US-08-482-142-88	Sequence 88, Appl
10	42	56.0	24	1	US-08-478-572-25	Sequence 25, Appl
11	42	56.0	24	1	US-08-478-572-88	Sequence 88, Appl

12	42	56.0	24	2	US-08-484-296-25	Sequence 25, Appl
13	42	56.0	24	2	US-08-484-296-88	Sequence 88, Appl
14	42	56.0	24	4	PCT-US95-04481-16	Sequence 16, Appl
15	42	56.0	104	2	US-09-755-100A-7	Sequence 7, Appli
16	42	56.0	183	2	US-09-755-100A-10	Sequence 10, Appl
17	42	56.0	218	2	US-09-755-100A-8	Sequence 8, Appli
18	42	56.0	218	2	US-09-755-100A-9	Sequence 9, Appli
19	42	56.0	222	1	US-07-945-288-11	Sequence 11, Appl
20	42	56.0	222	1	US-08-462-831-11	Sequence 11, Appl
21	42	56.0	222	1	US-08-461-809-11	Sequence 11, Appl
22	42	56.0	222	1	US-08-461-441-11	Sequence 11, Appl
23	42	56.0	222	4	PCT-US93-08518-11	Sequence 11, Appl
24	42	56.0	245	1	US-07-945-288-2	Sequence 2, Appli
25	42	56.0	245	1	US-08-462-831-2	Sequence 2, Appli
26	42	56.0	245	1	US-08-461-809-2	Sequence 2, Appli
27	42	56.0	245	1	US-08-461-441-2	Sequence 2, Appli
28	42	56.0	245	1	US-08-482-142-2	Sequence 2, Appli
29	42	56.0	245	1	US-08-478-572-2	Sequence 2, Appli
30	42	56.0	245	2	US-08-460-040-2	Sequence 2, Appli
31	42	56.0	245	2	US-08-484-296-2	Sequence 2, Appli
32	42	56.0	245	4	PCT-US93-08518-2	Sequence 2, Appli
33	42	56.0	320	1	US-07-945-288-10	Sequence 10, Appl
34	42	56.0	320	1	US-08-462-831-10	Sequence 10, Appl
35	42	56.0	320	1	US-08-461-809-10	Sequence 10, Appl
36	42	56.0	320	1	US-08-461-441-10	Sequence 10, Appl
37	42	56.0	320	4	PCT-US93-08518-10	Sequence 10, Appl
38	42	56.0	321	1	US-07-945-288-6	Sequence 6, Appli
39	42	56.0	321	1	US-08-462-831-6	Sequence 6, Appli
40	42	56.0	321	1	US-08-461-809-6	Sequence 6, Appli
41	42	56.0	321	1	US-08-461-441-6	Sequence 6, Appli
42	42	56.0	321	1	US-08-482-142-6	Sequence 6, Appli
43	42	56.0	321	1	US-08-478-572-6	Sequence 6, Appli
44	42	56.0	321	2	US-08-484-296-6	Sequence 6, Appli
45	42	56.0	321	4	PCT-US93-08518-6	Sequence 6, Appli
46	42	56.0	374	2	US-09-248-796A-19967	Sequence 19967, A
47	42	56.0	392	2	US-09-583-110-4374	Sequence 4374, Ap
48	42	56.0	399	2	US-09-107-433-3230	Sequence 3230, Ap
49	42	56.0	485	2	US-09-651-941-9	Sequence 9, Appli
50	42	56.0	485	2	US-09-955-597-9	Sequence 9, Appli
51	42	56.0	508	2	US-09-655-270A-9	Sequence 9, Appli
52	41.5	55.3	423	2	US-09-335-409-10	Sequence 10, Appl
53	41.5	55.3	423	2	US-09-568-102-10	Sequence 10, Appl
54	41.5	55.3	423	2	US-09-567-969-10	Sequence 10, Appl
55	41.5	55.3	423	2	US-09-568-480-10	Sequence 10, Appl
56	41.5	55.3	423	2	US-09-568-486-10	Sequence 10, Appl
57	41.5	55.3	423	2	US-09-568-472-10	Sequence 10, Appl
58	41.5	55.3	423	2	US-09-567-899-10	Sequence 10, Appl
59	41.5	55.3	423	2	US-10-014-717-10	Sequence 10, Appl
60	41	54.7	147	2	US-09-328-352-4890	Sequence 4890, Ap
61	41	54.7	150	2	US-10-101-464A-693	Sequence 693, App
62	41	54.7	162	2	US-09-575-574-4	Sequence 4, Appli
63	41	54.7	290	2	US-09-489-039A-8482	Sequence 8482, Ap
64	41	54.7	296	1	US-08-700-637-4	Sequence 4, Appli
65	41	54.7	300	2	US-09-902-540-16824	Sequence 16824, A
66	41	54.7	307	1	US-07-982-112-2	Sequence 2, Appli
67	41	54.7	354	2	US-09-489-039A-7697	Sequence 7697, Ap
68	41	54.7	478	2	US-09-605-703B-2160	Sequence 2160, Ap

```
 69    41    54.7     526   2   US-09-489-039A-10347      Sequence 10347, A
 70    41    54.7     705   2   US-09-252-991A-28353      Sequence 28353, A
 71    41    54.7     931   2   US-09-949-016-9850        Sequence 9850, Ap
 72    40    53.3      48   1   US-08-209-747-14          Sequence 14, Appl
 73    40    53.3      48   1   US-08-458-298-14          Sequence 14, Appl
 74    40    53.3     237   2   US-09-252-991A-20906      Sequence 20906, A
 75    40    53.3     248   2   US-09-640-211A-788        Sequence 788, App
 76    40    53.3     264   2   US-09-431-887-24          Sequence 24, Appl
 77    40    53.3     270   2   US-09-270-767-41818       Sequence 41818, A
 78    40    53.3     341   1   US-08-538-711A-8          Sequence 8, Appli
 79    40    53.3     341   2   US-08-725-027-8           Sequence 8, Appli
 80    40    53.3     341   2   US-09-542-552-8           Sequence 8, Appli
 81    40    53.3     353   1   US-08-538-711A-7          Sequence 7, Appli
 82    40    53.3     353   2   US-08-725-027-7           Sequence 7, Appli
 83    40    53.3     353   2   US-09-542-552-7           Sequence 7, Appli
 84    40    53.3     353   2   US-09-538-092-989         Sequence 989, App
 85    40    53.3     361   2   US-09-248-568-2           Sequence 2, Appli
 86    40    53.3     361   2   US-09-364-425B-19         Sequence 19, Appl
 87    40    53.3     361   2   US-09-364-425B-50         Sequence 50, Appl
 88    40    53.3     410   2   US-09-949-016-10345       Sequence 10345, A
 89    40    53.3     410   2   US-09-949-016-10346       Sequence 10346, A
 90    40    53.3     425   2   US-09-543-681A-5751       Sequence 5751, Ap
 91    40    53.3     821   2   US-09-252-991A-30347      Sequence 30347, A
 92    40    53.3    1841   1   US-08-804-227C-6          Sequence 6, Appli
 93    40    53.3    4630   2   US-09-091-609-2           Sequence 2, Appli
 94    40    53.3    5215   2   US-09-105-537-2           Sequence 2, Appli
 95    39.5  52.7     752   2   US-09-583-110-2714        Sequence 2714, Ap
 96    39.5  52.7     755   2   US-09-107-433-4628        Sequence 4628, Ap
 97    39.5  52.7    1088   2   US-09-130-242-2           Sequence 2, Appli
 98    39.5  52.7    1088   2   US-09-583-610D-2          Sequence 2, Appli
 99    39.5  52.7    1088   2   US-09-949-016-6935        Sequence 6935, Ap
100    39.5  52.7    1091   2   US-09-949-016-8595        Sequence 8595, Ap
```

ALIGNMENTS

RESULT 1
US-09-270-767-33522
; Sequence 33522, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33522
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33522

  Query Match   61.3%; Score 46; DB 2; Length 158;
  Best Local Similarity   70.0%; Pred. No. 31;

```
Matches    7;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        1 YGWGDGGYGS 10
            :||| ||| |
Db       78 FGWGYGGYAS 87


RESULT 7
US-10-164-595-2
; Sequence 2, Application US/10164595
; Patent No. 6657054
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies, Inc
;  TITLE OF INVENTION: Regulated Angiogenesis Genes and Polypeptides
;  FILE REFERENCE: 1U 103 R1
;  CURRENT APPLICATION NUMBER: US/10/164,595
;  CURRENT FILING DATE: 2002-06-10
;  NUMBER OF SEQ ID NOS: 80
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 378
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-164-595-2

  Query Match              56.7%;  Score 42.5;  DB 2;  Length 378;
  Best Local Similarity    64.3%;  Pred. No. 2.2e+02;
  Matches    9;  Conservative    0;  Mismatches    2;  Indels    3;  Gaps    1;

Qy        1 YGWGDGGY---GSD 11
            || |||||    | |
Db      260 YGGGDGGYNGFGGD 273


RESULT 8
US-08-482-142-25
; Sequence 25, Application US/08482142
; Patent No. 5820862
;  GENERAL INFORMATION:
;    APPLICANT:  Garman, Richard
;    APPLICANT:  Greenstein, Julia
;    APPLICANT:  Kuo, Mei-chang
;    APPLICANT:  Rogers, Bruce
;    APPLICANT:  Franzen, Henry
;    APPLICANT:  Chen, Xian
;    APPLICANT:  Evans, Sean
;    APPLICANT:  Shaked, Ze'ev
;    TITLE OF INVENTION:  T CELL EPITOPES OF THE MAJOR ALLERGENS
;    TITLE OF INVENTION:  FROM DERMATOPHAGOIDES (HOUSE DUST MITE)
;    NUMBER OF SEQUENCES:  207
;    CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  IMMULOGIC PHARMACEUTICAL CORPORATION
;     STREET:  610 LINCOLN STREET
;     CITY:  WALTHAM
;     STATE:  MA
;     COUNTRY:  USA
;     ZIP:  02154
```

```
;      COMPUTER READABLE FORM:
;        MEDIUM TYPE:  Floppy disk
;        COMPUTER:  IBM PC compatible
;        OPERATING SYSTEM:  PC-DOS/MS-DOS
;        SOFTWARE:  ASCII TEXT
;      CURRENT APPLICATION DATA:
;        APPLICATION NUMBER:  US/08/482,142
;        FILING DATE:  07-JUN-1995
;        CLASSIFICATION:  435
;      PRIOR APPLICATION DATA:
;        APPLICATION NUMBER:  US/08/445,307
;        FILING DATE:  07 June 1995
;      ATTORNEY/AGENT INFORMATION:
;        NAME:  CRAIG, ANNE I.
;        REGISTRATION NUMBER:  32,976
;        REFERENCE/DOCKET NUMBER:  017.6US
;      TELECOMMUNICATION INFORMATION:
;        TELEPHONE:  (617) 466-6000
;        TELEFAX:  (617) 466-6040
;    INFORMATION FOR SEQ ID NO:  25:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH:  24 amino acids
;        TYPE:  amino acid
;        TOPOLOGY:  linear
;      MOLECULE TYPE:  peptide
;      FRAGMENT TYPE:  N-terminal
US-08-482-142-25

    Query Match           56.0%;  Score 42;  DB 1;  Length 24;
    Best Local Similarity  85.7%;  Pred. No. 18;
    Matches    6;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps    0;

Qy          3 WGDGGYG 9
              ||| |||
Db         16 WGDNGYG 22


RESULT 14
PCT-US95-04481-16
; Sequence 16, Application PC/TUS9504481
;   GENERAL INFORMATION:
;     APPLICANT:
;     TITLE OF INVENTION:  Pharmaceutical Peptide Formulations For Treatment of
Dust Mit
;     NUMBER OF SEQUENCES:  54
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/04481
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/227,772
;       FILING DATE:  April 14, 1994
```

```
;     ATTORNEY/AGENT INFORMATION:
;        NAME:  Vanstone, Darlene A.
;        REGISTRATION NUMBER:  35,279
;        REFERENCE/DOCKET NUMBER:  017.5 PCT
;     TELECOMMUNICATION INFORMATION:
;        TELEPHONE:  (617) 466-6000
;        TELEFAX:  (617) 466-6040
;  INFORMATION FOR SEQ ID NO:  16:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH:  24 amino acids
;        TYPE:  amino acid
;        STRANDEDNESS:
;        TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     FRAGMENT TYPE:  internal
PCT-US95-04481-16
```

```
  Query Match            56.0%;  Score 42;  DB 4;  Length 24;
  Best Local Similarity  85.7%;  Pred. No. 18;
  Matches   6;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy        3 WGDGGYG 9
            ||| |||
Db       16 WGDNGYG 22
```

```
Search completed: December  8, 2005, 08:17:39
Job time : 38 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 8, 2005, 08:03:20 ; Search time 82.3636 Seconds
                                             (without alignments)
                                             64.015 Million cell updates/sec

Title: US-10-789-494B-5
Perfect score: 75
Sequence: 1 YGWGDGGYGSDS 12

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   A_Geneseq_21:*
                 1: geneseqp1980s:*
                 2: geneseqp1990s:*
                 3: geneseqp2000s:*
                 4: geneseqp2001s:*
                 5: geneseqp2002s:*
                 6: geneseqp2003as:*
                 7: geneseqp2003bs:*
                 8: geneseqp2004s:*
                 9: geneseqp2005s:*

        Pred. No. is the number of results predicted by chance to have a
        score greater than or equal to the score of the result being printed,
        and is derived by analysis of the total score distribution.

                                    SUMMARIES
                      %
Result   Query
   No.   Score   Match Length DB   ID   Description
----
       1   75   100.0   12   8   ADU51232   Adu51232 Gut silkw
       2   75   100.0   12   8   ADU51209   Adu51209 Silkworm
       3   75   100.0   14   8   ADU51180   Adu51180 Gut silkw
       4   75   100.0   21   8   ADU51182   Adu51182 Gut silkw
       5   75   100.0   21   8   ADU51178   Adu51178 Gut silkw
       6   75   100.0   21   8   ADU51192   Adu51192 Gut silkw
       7   75   100.0   21   8   ADU51185   Adu51185 Gut silkw
       8   75   100.0   22   8   ADU51201   Adu51201 Gut silkw

9	75	100.0	22	8	ADU51190	Adu51190	Gut silkw
10	75	100.0	22	8	ADU51188	Adu51188	Gut silkw
11	75	100.0	22	8	ADU51184	Adu51184	Gut silkw
12	75	100.0	45	8	ADU51173	Adu51173	Gut silkw
13	75	100.0	2655	7	ADO59401	Ado59401	Antheraea
14	71	94.7	21	8	ADU51183	Adu51183	Gut silkw
15	70	93.3	21	8	ADU51193	Adu51193	Gut silkw
16	69	92.0	21	8	ADU51194	Adu51194	Gut silkw
17	68	90.7	21	8	ADU51187	Adu51187	Gut silkw
18	63.5	84.7	25	8	ADU51196	Adu51196	Gut silkw
19	59	78.7	22	8	ADU51200	Adu51200	Gut silkw
20	55	73.3	23	8	ADU51176	Adu51176	Gut silkw
21	53	70.7	21	8	ADU51199	Adu51199	Gut silkw
22	51	68.0	170	8	ADT56783	Adt56783	Plant pol
23	49	65.3	124	5	ABB98794	Abb98794	Human cla
24	49	65.3	182	9	ADW17161	Adw17161	Eucalyptu
25	48	64.0	23	8	ADU51197	Adu51197	Gut silkw
26	48	64.0	129	8	ADY14170	Ady14170	Plant ful
27	48	64.0	214	9	ADY95170	Ady95170	Protein N
28	48	64.0	304	3	AAY81991	Aay81991	Tick alle
29	48	64.0	333	6	ABB80123	Abb80123	Blo t1. 6
30	48	64.0	346	6	ABB80124	Abb80124	Blo t1 fu
31	48	64.0	493	8	ADY13125	Ady13125	Plant ful
32	48	64.0	494	8	ADY12788	Ady12788	Plant ful
33	47	62.7	16	8	ADU51189	Adu51189	Gut silkw
34	47	62.7	17	8	ADU51174	Adu51174	Gut silkw
35	47	62.7	17	8	ADU51234	Adu51234	Gut silkw
36	47	62.7	806	4	AAE10035	Aae10035	N. mening
37	47	62.7	806	4	AAU27600	Aau27600	Neisseria
38	47	62.7	806	8	ADS00590	Ads00590	N. mening
39	47	62.7	998	8	ADS24441	Ads24441	Bacterial
40	46	61.3	67	4	AAU48381	Aau48381	Propionib
41	46	61.3	67	6	ABM44900	Abm44900	Propionib
42	46	61.3	145	4	AAU67654	Aau67654	Propionib
43	46	61.3	145	6	ABM64173	Abm64173	Propionib
44	46	61.3	146	5	ABP00239	Abp00239	Human ORF
45	46	61.3	190	4	AAU67377	Aau67377	Propionib
46	46	61.3	190	6	ABM63896	Abm63896	Propionib
47	46	61.3	1048	4	ABB64201	Abb64201	Drosophil
48	45	60.0	263	3	AAG36620	Aag36620	Arabidops
49	45	60.0	273	3	AAG36619	Aag36619	Arabidops
50	45	60.0	309	3	AAG36618	Aag36618	Arabidops
51	45	60.0	408	7	ADM26698	Adm26698	Hyperther
52	45	60.0	478	7	ADE28110	Ade28110	Human NTR
53	45	60.0	498	5	ABP27193	Abp27193	Streptoco
54	45	60.0	499	8	ADV81840	Adv81840	Streptoco
55	45	60.0	509	8	ADV88426	Adv88426	Streptoco
56	45	60.0	509	8	ADV79679	Adv79679	Streptoco
57	44.5	59.3	1103	8	ADT56583	Adt56583	Plant pol
58	44	58.7	96	3	AAG44861	Aag44861	Zea mays
59	44	58.7	141	4	ABB70805	Abb70805	Drosophil
60	44	58.7	265	3	AAG41623	Aag41623	Arabidops
61	44	58.7	302	8	ADY12065	Ady12065	Plant ful
62	44	58.7	304	3	AAG41622	Aag41622	Arabidops
63	44	58.7	306	7	ABM74277	Abm74277	DNA clone
64	44	58.7	318	4	AAE01958	Aae01958	Arabidops
65	44	58.7	318	5	AAU93117	Aau93117	Arabidops

```
 66     44    58.7     318   6   ADA15487              Ada15487 A. thalia
 67     44    58.7     318   6   ADB23126              Adb23126 Environme
 68     44    58.7     318   7   ADD30358              Add30358 Plant yie
 69     44    58.7     318   8   ADI43787              Adi43787 Plant tra
 70     44    58.7     318   8   ADO01683              Ado01683 Thalecres
 71     44    58.7     330   8   ADY04395              Ady04395 Plant ful
 72     44    58.7     345   7   ABM73644              Abm73644 DNA clone
 73     44    58.7     345   9   ADZ45295              Adz45295 Pineapple
 74     44    58.7     349   3   AAG35161              Aag35161 Zea mays
 75     44    58.7     353   3   AAG35160              Aag35160 Zea mays
 76     44    58.7     354   3   AAG41621              Aag41621 Arabidops
 77     44    58.7     354   3   AAG46149              Aag46149 Arabidops
 78     44    58.7     354   3   AAG18708              Aag18708 Arabidops
 79     44    58.7     386   7   ABM86123              Abm86123 Rice abio
 80     44    58.7     397   3   AAG35159              Aag35159 Zea mays
 81     44    58.7     434   9   AEB91457              Aeb91457 Microbial
 82     44    58.7     436   3   AAG46148              Aag46148 Arabidops
 83     44    58.7     436   3   AAG18707              Aag18707 Arabidops
 84     44    58.7     440   3   AAG46147              Aag46147 Arabidops
 85     44    58.7     440   3   AAG18706              Aag18706 Arabidops
 86     44    58.7     459   8   ADX78457              Adx78457 Plant ful
 87     44    58.7     492   8   ADY13411              Ady13411 Plant ful
 88     44    58.7     530   8   ADY09863              Ady09863 Plant ful
 89     44    58.7     534   8   ADT60817              Adt60817 Plant pol
 90     44    58.7     634   6   ADA54695              Ada54695 Human pro
 91     44    58.7    1170   8   ADN21199              Adn21199 Bacterial
 92     44    58.7    4861   5   AAU84280              Aau84280 Human end
 93     44    58.7    4861   6   AAE32729              Aae32729 HERC1 pro
 94     44    58.7    4861   7   ADC35083              Adc35083 Human bre
 95     44    58.7    4861   7   ADP65241              Adp65241 Human gua
 96     44    58.7    4899   4   ABB65885              Abb65885 Drosophil
 97    43.5   58.0     358   8   ADX96909              Adx96909 Plant ful
 98    43.5   58.0     364   8   ADX68278              Adx68278 Plant ful
 99    43.5   58.0     412   8   ADX72775              Adx72775 Plant ful
100     43    57.3      36   7   ADD26136              Add26136 Silkworm
```

<div align="center">ALIGNMENTS</div>

RESULT 1
ADU51232
ID   ADU51232 standard; peptide; 12 AA.
XX
AC   ADU51232;
XX
DT   24-FEB-2005 (first entry)
XX
DE   Gut silkworm fibroin peptide fragment 34.
XX
KW   vulnerary; cell proliferation; wound healing; cell adhesion; cosmetics;
KW   cell culture; fibroin.
XX
OS   Bombycoidea.
XX
PN   JP2004339189-A.
XX

PT   New peptide having excellent cell growth promoting activity, for use as a
PT   cell growth promoter, cell adhesion agent, wound healing-promoting agent,
PT   cosmetic and cell culture base material.
XX

PS   Example 3; Page; 27pp; Japanese.
XX

CC   The invention relates to a novel peptide having excellent cell growth
CC   promoting activity. The peptide of the invention demonstrates vulnerary
CC   activity and may be utilised as a cell growth promoter, cell adhesion
CC   agent, wound healing-promoting agent or cosmetic and cell culture base
CC   material. The current sequence is that of a gut silkworm fibroin peptide
CC   fragment of the invention which is described as being amorphous.
XX

SQ   Sequence 12 AA;

      Query Match   100.0%;   Score 75;   DB 8;   Length 12;
      Best Local Similarity   100.0%;   Pred. No. 0.0011;
      Matches   12; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 YGWGDGGYGSDS 12
             ||||||||||||
Db   1 YGWGDGGYGSDS 12

RESULT 22
ADT56783
ID   ADT56783 standard; protein; 170 AA.
XX

AC   ADT56783;
XX

DT   13-JAN-2005   (first entry)
XX

DE   Plant polypeptide, SEQ ID 6860.
XX

KW   Plant; transgenic; cold tolerance; growth rate; drought tolerance;
KW   disease resistance; galactomannan production; plant growth regulator;
KW   heat tolerance; herbicide tolerance; lignin production;
KW   extreme osmotic condition tolerance; pathogens resistance;
KW   pest resistance; yield improvement; seed oil yield; seed protein yield.
XX

OS   Viridiplantae.
XX

PN   US2004216190-A1.
XX

PD   28-OCT-2004.
XX

CC   The invention relates a recombinant DNA construct comprising a
CC   polynucleotide having any of 5544 nucleotide sequences (cDNAs SEQ ID NO:
CC   1-5544) and encoding a polypeptide with any of 5544 amino acid sequences
CC   (SEQ ID NO: 5545-11088). The cDNAs and proteins are from corn, soybean,
CC   Arabidopsis, wheat and rape but the specification does not indicate which
CC   sequences is derived from which organism. Also included is a method of
CC   producing a plant having an improved property, comprising transforming a
CC   plant with a recombinant DNA construct comprising a promoter region
CC   functional in a plant cell operably joined to a polynucleotide encoding a
CC   polypeptide associated with the property, and growing the transformed
CC   plant. The property is selected from improving plant cold tolerance, for
CC   manipulating growth rate in plant cells by modification of the cell cycle
CC   pathway, for improving plant drought tolerance, for providing increased
CC   resistance to plant disease, for galactomannan production, for production
CC   of plant growth regulators, for improving plant heat tolerance, for
CC   improving plant tolerance to herbicides, for increasing the rate of
CC   homologous recombination in plants, for lignin production, for improving
CC   plant tolerance to extreme osmotic conditions, for improving plant
CC   tolerance to pathogens or pests, for yield improvement by modification of
CC   photosynthesis, for modifying seed oil yield and/or content, for
CC   modifying seed protein yield and/or content, for yield improvement by
CC   modification of carbohydrate, nitrogen or phosphorus use and/or uptake
CC   and for yield improvement by providing improved plant growth and
CC   development under at least one stress condition. The polynucleotide may
CC   also encode a plant transcription factor. The methods and compositions of
CC   the present invention are useful in the field of biochemistry and
CC   genetics, in particular for producing transgenic plants with improved
CC   biological characteristics such as increased yield, improved nitrogen
CC   flow, increasing plant tolerance to cold or heat, improving plant
CC   tolerance to extreme osmotic and drought conditions, and improving plant
CC   tolerance to plant pests or pathogens. They can also be used in physical
CC   arrays of molecules, plant breeding markers, computer-based storage and
CC   analysis systems. The present sequence is one of the 5544 plant protein
CC   sequences of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20040216190.
XX
SQ   Sequence 170 AA;

```
Query Match               68.0%;   Score 51;   DB 8;   Length 170;
Best Local Similarity     75.0%;   Pred. No. 25;
Matches     9;   Conservative     0;   Mismatches     3;   Indels     0;   Gaps     0;

Qy          1 YGWGDGGYGSDS 12
              || ||||||| |
Db        154 YGGGDGGYGGGS 165


Search completed: December  8, 2005, 08:10:41
Job time : 86.3636 secs
```